PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

AM

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/12, 15/62, A61K 37/02 — A1

(11) International Publication Number: WO 94/14958

(43) International Publication Date: 7 July 1994 (07.07.94)

(21) International Application Number: PCT/AU93/00655

(22) International Filing Date: 16 December 1993 (16.12.93)

(30) Priority Data:

PL 6520	22 December 1992 (22.12.92)	AU
PL 9661	28 June 1993 (28.06.93)	AU

(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, LV, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(71) Applicant *(for all designated States except US)*: THE UNIVERSITY OF SYDNEY [AU/AU]; Parramatta Road, Sydney, NSW 2006 (AU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WEISS, Anthony, Steven [AU/AU]; 235 Rainbow Street, Randwick, NSW 2031 (AU). MARTIN, Stephen, Lewis [GB/GB]; Kewstoke, Greenleighs, Sedgley, West Midlands DY3 3RZ (GB).

(74) Agent: GRIFFITH HACK & CO.; GPO Box 4164, Sydney, NSW 2001 (AU).

(54) Title: SYNTHETIC POLYNUCLEOTIDES

(57) Abstract

This invention provides recombinant tropoelastins and variants of recombinant tropoelastins produced from synthetic polynucleotides, as well as the synthetic polynucleotides themselves. The invention also provides cross-linked elastins or elastin-like products prepared from the tropoelastins or variants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

SYNTHETIC POLYNUCLEOTIDES

TECHNICAL FIELD

The present invention relates to the production of recombinant tropoelastins, and variants of these recombinant tropoelastins, from synthetic polynucleotides, and uses of the tropoelastins and variants.

BACKGROUND ART

There are various forms of tropoelastin that typically appear to consist of two types of alternating domains: those rich in hydrophobic amino acids (responsible for the elastic properties) and those rich in lysine residues (responsible for cross-link formation). Hydrophobic and cross-linking domains are encoded in separate exons (Indik *et al.*, 1987).

The gene for tropoelastin is believed to be present as a single copy in the mammalian genome, and is expressed in the form of multiple transcripts, distinguished by alternative splicing of the pre-mRNA (Indik *et al*, 1990; Oliver *et al*, 1987).

Previous recombinant work with tropoelastin has been reported by Indik *et al* (1990) who achieved modest expression of a natural human tropoelastin sequence from cDNA. Their product was unstable, the free polypeptide being rapidly degraded.

Bressan *et al* (1987) have reported the cloning of a defined naturally occurring segment of chick tropoelastin.

DESCRIPTION OF THE INVENTION

The present invention provides for the expression of significant amounts of tropoelastins or variants of the tropoelastins in recombinant expression systems.

The present inventors have recognised that tropoelastins are proteins which can be used in a variety of, for instance, pharmaceutical applications, but these uses require significant quantities of tropoelastin. These quantities could be obtained by cloning naturally occurring tropoelastin genes, but the present inventors show how they can be more easily obtained by producing

synthetic polynucleotides adapted to provide enhanced expression.

The present inventors have recognised that because tropoelastins have highly repetitive coding sequences, the tropoelastin genes have the potential to include significant numbers of codons which have low usage in particular hosts. Codons of low usage can hamper gene expression.

For example, in one tropoelastin coding sequence described in detail in this application, the natural sequence contains of the order of 80 glycine GGA codons which comprises 10% of the gene and have low usage in *Escherichia coli* [Fazio *et al.*, 1988, and Genetics Computer Group (GCG) package version 7-UNIX using Codon Frequency and Gen Run Data: ecohigh-cod].

According to a first aspect of the present invention, there is provided a synthetic polynucleotide encoding the amino acid sequence of a tropoelastin or a variant of the tropoelastin.

The tropoelastin may be a mammalian or avian tropoelastin such as human, bovine, ovine, porcine, rat or chick tropoelastin. Preferably, the tropoelastin is human tropoelastin.

The synthetic polynucleotide sequence is altered with respect to the natural coding sequence for the tropoelastin molecule or variant so that:

a) it codes for a tropoelastin sequence or a variant of the tropoelastin; and

b) all or some of the codons which hamper expression in the expression system in which the polynucleotide is to be expressed, are replaced with codons more favourable for expression in the expression system.

Preferably all, or part, of the 5' or 3' untranslated regions, or both, of the natural coding sequence are excluded from the synthetic polynucleotide.

Preferably all, or part, of the signal peptide encoding region is excluded from the synthetic polynucleotide.

Where the synthetic polynucleotide is prepared from assembled oligonucleotides it is preferred to incorporate restriction sites in the sequence to facilitate assembly of the polynucleotide.

Restriction sites incorporated in the polynucleotide sequence are also useful for:

1. facilitating subcloning of manageable blocks for sequence confirmation;

2. providing sites for later introduction of modifications to the polynucleotide as insertions, deletions or base changes;

3. facilitating confirmation of correct polynucleotide assembly by restriction endonuclease digestion.

A preferred expression system is an *Escherichia coli* expression system. However, the invention includes within its scope synthetic polynucleotides suitable for use in other expression systems such as other microbial expression systems. These other expression systems include yeast and bacterial expression systems, insect cell expression systems, and expression systems involving other eukaryotic cell lines or whole organisms.

Modifications to codon usage to provide enhanced expression are discussed in:

Zhang *et al* (1991) for *E. coli*, yeast, fruit fly and primates where codon usage tables are provided;

Newgard *et al* (1986) for mammals; and Murray *et al* (1989) for plants. Preferred codon usages are indicated in these publications.

Preferably, at least 50% of codons for any particular amino acid are selected and altered to reflect preferred codon usage in the host of choice.

Preferably, the polynucleotide is a fused polynucleotide with the tropoelastin or variant encoding sequence fused to a polynucleotide sequence compatible with the host. The compatible sequence is preferably at the 5' end of the polynucleotide molecule.

Preferred compatible polynucleotides include those

which encode all or part of a polypeptide which causes the expressed fusion to be secreted or expressed as a cell surface protein so as to facilitate purification of the expressed product, or expressed as a cytoplasmic protein.

One preferred compatible polynucleotide is one encoding all or part of glutathione-S-transferase.

In addition the synthetic polynucleotides can encode additional residues such as an N-terminal methionine or f-methionine not present in the natural counterpart.

A preferred synthetic polynucleotide is one comprising the sequence illustrated in Figure 3 (1) to 3 (5) (SEQ ID NO 1) or a part of it, encoding a polypeptide which retains elastic properties. The sequence illustrated in Figure 3 (1) to 3 (5) is 2210 bp in size.

To our knowledge, this is the largest synthetic gene constructed so far. Previously, the largest was of the order of 1.5 kb in size.

The actual changes made in this sequence in comparison with the natural sequence from which it was derived are shown in Figure 6 (1) to 6 (4) comparing the synthetic sequence (SEQ ID NO 1) with the natural sequence (SEQ ID NO 53). Synthetic polynucleotides in which only some of the base changes shown in that Figure have been made are also within the scope of the invention.

It is known that tropoelastin genes in nature are expressed as multiple transcripts which are distinguished by alternative splicing of the pre-mRNA as described in, for instance:

Indik *et al*, 1990; Oliver *et al*, 1987; Heim *et al*, 1991; Raju *et al*, 1987; and Yeh *et al*, 1987. The tropoelastins of the present invention for which synthetic polynucleotides are prepared are intended to encompass these different splice forms.

Variants of tropoelastins embodying the present invention are polypeptides which retain the basic structural attributes, namely the elastic properties, of

a tropoelastin molecule, and which are homologous to naturally occurring tropoelastin molecules. For the purposes of this description, "homology" between two sequences connotes a likeness short of identity indicative of a derivation of one sequence from the other. In particular, a polypeptide is homologous to a tropoelastin molecule if a comparison of amino-acid sequences between the molecules reveals an identity of greater than about 65% over any contiguous 20 amino acid stretch or over any repetitive element of the tropoelastin molecule shorter than 20 amino acids in length. Such a sequence comparison can be performed via known algorithms, such as the one described by Lipman and Pearson, Science 227 : 1435 (1985) which are readily implemented by computer.

Variants of tropoelastins can be produced by conventional site-directed or random mutagenesis. This is one avenue for routinely identifying residues of the molecule that can be modified without destroying the elastic properties of the molecule.

Oligonucleotide-directed mutagenesis, comprising:

1. synthesis of an oligonucleotide with a sequence that contains the desired nucleotide substitution (mutation),

2. hybridizing the oligonucleotide to a template comprising a structural sequence coding for tropoelastin and

3. using a DNA polymerase to extend the oligonucleotide as a primer, is preferred because of its ready utility in determining the effects of particular changes to the structural sequence. Its relative expense may militate in favour of an alternative, known direct or random mutagenesis method.

Another approach which is particularly suited to situations where the synthetic polynucleotide has been prepared from oligonucleotide blocks bounded by restrictions sites is cassette mutagenesis where entire restriction fragments are inserted, deleted or replaced.

Also exemplary of variants within the present

invention are molecules that correspond to a portion of a tropoelastin molecule without being coincident with a natural tropoelastin molecule and which retain the elastic properties of a natural tropoelastin molecule.

Other variants of tropoelastins of the present invention are fragments that retain the elastic properties of a tropoelastin molecule.

Fragments within the scope of this invention are typically greater than 20 amino acids in length.

According to a second aspect of the present invention there is provided a recombinant DNA molecule comprising a synthetic polynucleotide of the first aspect, and vector DNA.

Vectors useful in the invention include plasmids, phages and phagemids. The synthetic polynucleotides of the present invention can also be used in integrative expression systems or lytic or comparable expression systems.

Suitable vectors will generally contain origins of replication and control sequences which are derived from species compatible with the intended expression host. Typically these vectors include a promoter located upstream from the synthetic polynucleotide, together with a ribosome binding site for prokaryotic expression, and a phenotypic selection gene such as one conferring antibiotic resistance or supplying an auxotrophic requirement. For production vectors, vectors which provide for enhanced stability through partitioning may be chosen. Where integrative vectors are used it is not necessary for the vector to have an origin of replication. Lytic and other comparable expression systems do not need to have those functions required for maintenance of vectors in hosts.

Typical vectors include pBR322, pBluescript II SK^+, pGEX-2T, pTrc99A, pET series vectors, particularly pET3d, (Studier et al; 1990) and derivatives of these vectors.

According to a third aspect of the present invention there is provided a transformed host transformed with a

recombinant DNA molecule of the second aspect.

Hosts embodying the invention include bacteria, yeasts, insect cells and other eukaryotic cells or whole organisms. They are typically bacterial hosts.

A preferred host is an *E. coli* strain. Examples of *E. coli* hosts include *E. coli* B strain derivatives (Studier *et al*, 1990), NM522 (Gough and Murray, 1983) and XL1-Blue (Bullock *et al*, 1987). Hosts embodying this invention, for providing enhanced expression of tropoelastin or tropoelastin variants, are those in which the altered codon usage is favourable for expression, and with which any control sequences present in the recombinant DNA are compatible.

According to a fourth aspect of the present invention there is provided an expression product of a transformed host of the third aspect which expression product comprises a tropoelastin or a variant thereof.

A preferred expression product of the fourth aspect comprises all or part of the amino-acid sequence depicted in Figure 3 (1) to 3 (5) (SEQ ID NO: 1). The serine at position 1 may be deleted from the product and similarly the methionine at position 2 may be deleted.

Other preferred expression products are those in which only some of the base changes shown in Figure 6 (1) to 6 (4) have been made. Typically at least 50% of the indicated base changes have been made.

The expression products of the fourth aspect may be fused expression products which include all or part of a protein encoded by the vector in peptide linkage with the expression product. They may also include, for example, an N-terminal methionine or other additional residues which do not impair the elastic properties of the product.

Typically the fusion is to the N-terminus of the expression product. An example of a suitable protein is glutathione-S-transferase. The fused protein sequence may be chosen in order to cause the expression product to be secreted or expressed as a cell surface protein to

simplify purification or expressed as a cytoplasmic protein.

The expressed fusion products may subsequently be treated to remove the fused protein sequences to provide free tropoelastin or a free tropoelastin variant.

The expression products of the fourth aspect may also be produced from non-fusion vectors such as pND211 (N. Dixon, Australian National University). This vector has the gene inserted into an NcoI site and uses lambda-promoter-driven expression to permit initiation from the start codon of the synthetic gene. The sequence of the vector is shown at Figure 9 (1) and 9 (2) (SEQ ID NO: 54). Other suitable non-fusion vectors include pET3d.

According to a fifth aspect of the present invention there is provided a pharmaceutical or veterinary composition comprising an expression product of the fourth aspect together with a pharmaceutically or veterinarally acceptable carrier.

Dosage of the expression product and choice of carrier will vary with the specific purpose for which the expression product is being administered.

The expression products of the fourth aspect may also be prepared in the form of foods or as industrial products where elastic or association properties may be desired. The tropoelastin expression products of the invention can form associations in solution wherein the tropoelastin molecules are held together by hydrophobic interactions. These associations are termed "coacervates". They are useful as precursors to elastin synthesis. The tropoelastin coacervates can also be used as delivery vehicles for active ingredients such as pharmaceutical or veterinary agents providing biodegradable or biodissociable slow release formulations or alternatively protective coatings to protect active agents, for instance, during their transit through the stomach of a host.

According to a sixth aspect of the present invention there is provided a process for the production of an

expression product of the fourth aspect comprising:

providing a transformed host of the third aspect; culturing it under conditions suitable for the expression of the product of the fourth aspect; and collecting the expression product.

In one preferred form the expression product is produced in the form of inclusion bodies which are harvested from the transformed host.

In a seventh aspect of the invention there is provided a cross-linked expression product of the fourth aspect. The cross-linked expression products form elastin or elastin-like products.

In preparing a synthetic polynucleotide in accordance with the first aspect the following procedure is followed.

A cDNA sequence encoding a tropoelastin, or a part of it, is selected and the open reading frame is defined.

The sequence is then translated to provide the corresponding amino acid sequence. Alternatively, the procedure can commence from a known amino acid sequence.

The exons which are to be included in the expression product are chosen. Preferably, any signal sequence or untranslated regions will not be included in the synthetic polynucleotide.

The amino acid sequence selected is then converted to a polynucleotide sequence on the basis of codon usage frequencies. By selecting the most commonly used codon for each amino acid for the host in which expression is desired, a skewed usage arises because particular codons may have very different frequencies of usage. It is therefore necessary to adjust the codon usage of at least the most common codons, that is, those present at greater than 20 occurrences, to more closely match levels of codon usage in the host of choice.

It is preferable to alter the sequence to introduce restriction sites at regular intervals throughout the sequence where these represent silent alterations, that is, they do not change the resulting amino acid. In

addition ends suitable for ligation, eg BamHI and/or NcoI sites can be introduced into the sequence.

Tropoelastin sequences described for various organisms are similar, particularly at the level of exon structure and the organisation of hydrophilic and hydrophobic domains. In selecting exons to be included in the expression product we have adopted an approach whereby we leave in exons known to occur in all available tropoelastins. Depending on the intended use of the resulting tropoelastin, additional exons, or synthetic sequences, or both, are included. For instance, in the human example provided we included exon 10A which only occurs in some of the known sequences for human tropoelastin. In the bovine case, a typical addition would be exons 4A, 6 and/or 9 (Raju and Anwar, 1987; Yeh et al, 1987). In the rat case, a typical addition would be exons corresponding to exons 12 through 15 of the bovine case. (Heim et al 1991).

The construction of the synthetic polynucleotide of Figures 3 and 6 will now be described in more detail.

The synthetic tropoelastin gene described here differs from the natural coding sequence(s) in a number of ways. The untranslated regions present in the tropoelastin cDNA sequence were disregarded in designing the synthetic gene, and the nucleotides encoding the signal peptide were removed. Restriction endonuclease recognition sites were incorporated at regular intervals into the gene by typically altering only the third base of the relevant codons, thereby maintaining the primary sequence of the gene product. The facility for silent alteration of the coding sequence was also exploited to change the codon bias of the tropoelastin gene to that commonly found in highly expressed *E.coli* genes. [Genetics Computer Group (GCG) package version 7-UNIX using Codon Frequency and Gen Run Data: ecohigh-cod]. Two additional stop codons were added to the 3'-end, and an ATG start codon comprising a novel NcoI site was appended to the 5'-end. *Bam* HI cloning sites were

engineered at both ends of the synthetic sequence. Since the gene contains no internal methionine residues, treatment of the newly-synthesized gene product (expressed directly or as a fusion with another gene) with cyanogen bromide would liberate a protein with the same or similar sequence as one form of natural tropoelastin comprising 731 amino acids. Other forms of processing are envisaged, which may generate tropoelastin species of the same or different lengths.

Two stop codons were added in order to allow the possible use of the construct in suppressor hosts, and also to avoid any potential depletion of termination (release) factors for translation.

The inclusion of an ATG site is useful because: (1) it provides an appropriate restriction site for cloning, although this is a flexible property; (2) it provides a potential start codon for translation of an unfused synthetic gene; and (3) it introduces a methionine which can be cleaved by cyanogen bromide to release the tropoelastin species. However, another method of cleavage would not necessarily rely upon the availability of this methionine.

Fusion can provide a more stably expressed protein, and experience of other workers has suggested that unfused tropoelastin may be unstable (Indik *et al.*, 1990). The fusion is typically to the carboxy terminus of the fusion protein (i.e. the N-terminus of the tropoelastin). Glutathione-S-transferase (Smith and Johnson, 1988) is an example of a suitable fusion protein.

A convergent approach was used in assembly and cloning of the synthetic human tropoelastin (SHEL) sequence. Groups of six, and in one case, eight, oligonucleotides were annealed and ligated together to generate eight synthetic blocks of approximately 260-300bp, designated SHEL1-8. These blocks were cloned independently into pBluescript II SK$^+$; the assembly and cloning scheme for SHEL1 is illustrated in Figure 1.

Following sequence confirmation, the blocks were excised from their parent plasmids and used to construct three clones, pSHEL α,ß and γ, each containing approximately 700-800bp of the synthetic gene. The final step towards assembly of the complete SHEL gene involved ligation of the inserts from each of these three intermediary clones into pBluescript II SK$^+$ to produce pSHEL. The cloning scheme is illustrated in Figure 2.

The tropoelastin or variant produced as an expression product from vectors such as pSHEL can be chemically cross-linked to form an elastin product. Three available procedures are:

1. chemical oxidation of lysine side chains which are conducive to cross-linking [eg ruthenium tetroxide-mediated oxidation, via the amide (Yoshifuji S; Tanaka K; and Nitto Y (1987) Chem. Pharm Bull 35 2994-3000) and quinone-mediated oxidation];

2. homobifunctional chemical cross-linking agents, such as dithiobis(succinimidylpropionate), dimethyl adipimidate and dimethyl pimelimidate. There are many other amine-reactive cross-linking agents which could be used as alternatives; and

3. cross-linking via lysine and glutamic acid side chains as taught by Rapaka et al (1983).

The tropoelastins or variants of the invention may also be enzymatically cross-linked to form an elastin or elastin-like product. Enzymatic methods include lysyl oxidase-mediated oxidation of the tropoelastin or variant via modification of peptidyl lysine [Beddell-Hogan et al (1993)]. Oxidised lysines participate in the generation of cross-linkages between and within tropoelastin molecules. Other modification enzymes can be used forming cross-links via lysine or other residues.

Cross-linking can also be achieved by gamma irradiation using, for instance, techniques adapted from Urry et al (1986).

Tropoelastins or variants of the invention cross-linked to form elastin or elastin-like products are also

within the scope of the invention.

The half-lives of the products in free solution will determine the suitability of a particular agent for a particular application.

For example, the hydrolytic breakdown of the cross-linked material will be useful in applications, such as surgical applications, where the gradual loss of material over time is intended.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention is further described with reference to the accompanying drawings in which:

Figure 1 shows the scheme for construction and cloning of SHEL1, one of the eight intermediary subassemblies used to generate the SHEL sequence. A similar approach was adopted for each of the remaining blocks (sHEL 2-8). See materials and methods section for details. 5'-phosphorylated oligonucleotides are indicated with a black dot (•).

Figure 2 shows the cloning scheme for the synthetic human tropoelastin (SHEL). - **Abbreviations**: B,*Bam* HI; H,*Hind*III; K,*Kpn*I; N,*Not*I; P,*Pst*I; S,*Sac*I; Sp,*Spe*I.

Figure 3 (1) to 3 (5) shows over 5 drawing sheets the full nucleotide sequence (SEQ ID NO: 1) and corresponding amino acid sequence (SEQ ID NO: 2) for the synthetic human tropoelastin (SHEL). Coding (+) strand of the sHEL gene construct is shown on the upper (numbered) sequence line. Synthetic complementary (-) strand sequence is shown immediately beneath it. The amino acid sequence of the synthetic gene product is indicated below the nucleotide sequence.

Figure 4 (1) to 4 (2) shows over 2 drawings sheets the sequences for the oligonucleotides (SEQ ID NOS: 3 to 27) used to construct the synthetic human tropoelastin (SHEL) sequence: (+)- strand oligonucleotides.

Figure 5 (1) to 5 (2) shows over 2 drawing sheets the sequences for the oligonucleotides (SEQ ID NOS: 28 to 52) used to construct the synthetic human tropoelastin

(SHEL) sequence: (-) - strand oligonucleotides.

Figure 6 (1) to 6 (4) shows over 4 drawing sheets the differences in nucleotide sequence between SHEL (SEQ ID NO: 1) and a cDNA form of the coding region of the human tropoelastin gene (SEQ ID NO: 53). The coding (+)-strand of the synthetic (SHEL) sequence is shown on the top (numbered line). The cDNA sequence is indicated below it, showing only those nucleotides which differ from the synthetic sequence.

Figure 7 shows the results of SDS-PAGE analysis of tropoelastin fusion protein expression from pSHELC. Lane 1: standards; Lane 2: non-induced; Lane 3: induced. The arrow points to the overexpressed fusion protein.

Figure 8 shows the correlation between predicted and observed amino acid content for the fusion protein expressed from pSHELC: —Δ— Net data (%)
--O-- Expected (%)

Figure 9 (1) to 9 (2) over 2 drawing sheets shows the sequence (SEQ ID NO: 54) of the plasmid vector pND211.

Figure 10 shows the results of SDS-PAGE analysis of tropoelastin expression from pSHELF.

Lane 1: standards; Lane 2: induced; Lane 3: uninduced: Lane 4: alcohol-purified sample; Lane 5: additional lane of alcohol purified sample.

Figure 11 shows the correlation between predicted and observed amino acid content for tropoelastin expressed from pSHELF.

BEST METHOD OF PERFORMING THE INVENTION

The recombinant and synthetic techniques used are standard techniques which are described in standard texts such as Sambrook *et al* (1989).

Purification of the expression products is also performed using standard techniques, with the actual sequence of steps in each instance being governed by the host/expression product combination.

The pharmaceutical and veterinary compositions are formulated in accordance with standard techniques.

The amount of expression product that may be combined with carrier to produce a single dosage form will vary depending upon the condition being treated, the host to be treated and the particular mode of administration.

It will be understood, also, that the specific dose level for any particular host will depend upon a variety of factors including the activity of the expression product employed, the age, body weight, general health, sex, diet of the patient, time of administration, route of administration, rate of excretion, drug combination, etc.

The compositions may be administered parenterally in dosage unit formulations containing conventional, non-toxic, pharmaceutically and/or veterinarally acceptable carriers, diluents, adjuvants and/or excipients as desired.

Injectable preparations, for example, sterile injectable aqueous or oleagenous suspensions may be formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution or suspension in a non-toxic parenterally acceptable diluent or solvent. Among the acceptable vehicles or solvents that may be employed are water, Ringer's solution, and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil may be employed including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid and organic solvents find use in the preparation of injectables.

Routes of administration, dosages to be administered as well as frequency of administration are all factors which can be optimised using ordinary skill in the art.

In addition, the expression products may be prepared

as topical preparations for instance as anti-wrinkle and hand lotions using standard techniques for the preparation of such formulations. They may be prepared in aerosol form for, for instance, administration to a patient's lungs, or in the form of surgical implants, foods or industrial products by standard techniques.

The tropoelastins can be cross-linked either chemically, enzymatically or by irradiation to form elastin products for use in applications such as pharmaceutical applications, surgical, veterinary and medical applications, cosmetic applications, and in industrial uses. Tropoelastin coacervates can be used to formulate slow release compositions of active ingredients or to form protective coatings for active ingredients using standard formulation techniques.

Materials and Methods

Materials

Restriction enzymes, T4 polynucleotide kinase and T4 DNA ligase were obtained from Boehringer Mannheim, Progen Industries or New England Biolabs. Gelase® was obtained from Epicentre Technologies. Reagents for solid-phase oligodeoxynucleotide synthesis were obtained from Applied Biosystems (ABI). Low melting temperature (LMT) agarose was obtained from Progen or FMC and α-^{35}S-dATP was obtained from Amersham International. Plasmid vectors pBluescript II SK$^+$ and pGEX-2T were obtained from Stratagene and Medos Co Pty Ltd respectively. pET3d was obtained from F.W. Studier at Brookhaven National Laboratory, NY, U.S.A. E. coli strains HMS174 and BL21 (DE3) are described in Studier *et al* (1990).

Oligodeoxynucleotide Synthesis and Purification

Oligonucleotides were synthesized on 40nmol-scale polystyrene-support columns on an Applied Biosystems 381A or 394 DNA synthesis machine. Standard ABI protocols were employed for synthesis, including chemical 5'-phosphorylation where appropriate. Detritylation was performed automatically, and cleavage from the solid support effected manually (381A) or automatically (394)

according to the synthesizer used. Base protecting groups were removed by heating the ammoniacal oligonucleotide solution at 55-60°C overnight. Deprotected oligonucleotides were lyophilized, dissolved in 400μl TE buffer and ethanol precipitated prior to resuspension in 100μl 50% deionized formamide in TE.

All oligonucleotides used in construction of the sHEL gene were purified by denaturing PAGE before use. 160mm x 100mm x 1.5mm polyacrylamide gels containing 7M urea were used for this purpose. Short oligonucleotides (<40-mers) were purified on 20% gels whilst long oligonucleotides (>85-mers) were purified on gels containing 8-10% acrylamide (acrylamide:bisacrylamide 19:1). Samples were heated to 75°C for 3 min before loading. Tracking dye (0.05% bromophenol blue, 0.05% xylene cyanole FF in deionized formamide) was loaded into an adjacent lane. Electrophoresis was conducted at constant power (17W) until the bromophenol blue marker was within 1cm of the base of the gel. The apparatus was disassembled and the gel wrapped in cling film. Product bands were visualized by UV-shadowing over a fluorescent TLC plate. Excised gel fragments containing purified oligonucleotides were transferred to microcentrifuge tubes, crushed and soaked overnight at 60°C in 500μl elution buffer (0.3M sodium acetate pH7.0). A second extraction was performed with 400μl elution buffer, for 3-4h at 60°C and the supernatant combined with that of the first extraction. The total volume of the oligonucleotide-containing solution was reduced to approximately 400μl by butan-1-ol extraction and DNA precipitated by addition of 1ml ethanol. Purified oligonucleotide was pelleted by centrifugation, redissolved in 20μl TE buffer and quantified by spectrophotometry. The final yield of purified oligonucleotide obtained in this manner was typically 10-30μg.

Construction of Synthetic Gene 'Blocks' (sHEL1-8)

Complementary oligonucleotides (30pmol each, approx

1μg for 95-mers) were annealed in 10μl buffer containing 50mM Tris.HCl pH7.5, 10mM $MgCl_2$. The mixture was overlayed with 12μl paraffin oil, heated to 95°C and cooled slowly to 16°C (16h) in a microprocessor-controlled heating block (Perkin Elmer Cetus Thermal Cycler). Annealed samples were transferred to clean microcentrifuge tubes and a small aliquot (1μl) withdrawn for analysis by agarose gel electrophoresis (2%LMT gel, TBE running buffer). For each block comprising three complementary oligonucleotide pairs, four separate ligation reactions were set up. Each contained 50mM Tris.HCl pH7.5, 10mM $MgCl_2$, 1mM ATP, 3mM DTT, 3μl each of the appropriate annealed samples, 0.5μl (0.5U) T4 DNA ligase and Milli-Q water to a total volume of 10μl. All components except the ATP, DTT and T4 ligase were mixed and heated to 55°C for 5 min to denature cohesive termini and cooled to room temperature before addition of the remaining components. Ligation reactions were incubated overnight at 16°C and analysed on 2% LMT agarose gels, with TBE as running buffer. Ligated blocks were purified by preparative agarose gel electrophoresis using 2% LMT agarose gels with TAE running buffer. Product bands were identified under long-wave UV illumination with reference to known DNA size standards (pBluescript II SK^+ digested with *Hae* III) and excised in the minimum possible volume of gel. DNA was recovered from LMT agarose fragments using Gelase® in accordance with the manufacturer's instructions ("fast" protocol). Purity and yield of recovered sHEL blocks was assessed by analytical agarose electrophoresis alongside known DNA size standards. Block 8 was created by a slightly different strategy. The first 3 oligonucleotide pairs (numbers 22, 23, 24, 47, 48 and 49) were assembled and purified as described for blocks 1 to 7, after which the remaining oligonucleotide pair (numbers 25 and 50) was ligated under conditions described above. The full length block 8 was purified as described for blocks 1 to 7.

The oligonucleotides used for preparing each of the blocks shown in Figures 4 (1) to 4 (2) and 5 (1) to 5 (2) were assembled as follows:

Block	+strand oligonucleotides	Seq ID	-strand oligonucleotides	Seq ID
1	1,2,3	3 - 5	26,27,28	28 - 30
2	4,5,6	6 - 8	29,30,31	31 - 33
3	7,8,9	9 - 11	32,33,34	34 - 36
4	10,11,12	12 - 14	35,36,37	37 - 39
5	13,14,15	15 - 17	38,39,40	40 - 42
6	16,17,18	18 - 20	41,42,43	43 - 45
7	19,20,21	21 - 23	44,45,46	46 - 48
8	22,23,24,25	24 - 27	47,48,49,50	49 - 52

Blocks 1-8: Cloning

pBluescript II SK$^+$ DNA was digested with appropriate restriction enzymes and purified at each stage by preparative gel electrophoresis (1% agarose, TAE buffer). Plasmid DNA was isolated from agarose using a proprietary DNA purification matrix (Prep-A-Gene,Bio-Rad). Approximately 100ng (*ca.* 0.05pmol) of purified plasmid fragment was added to 50ng (*ca.* 0.3pmol) synthetic block in 17μl buffer containing 50mM Tris.HCl pH7.5, 10mM $MgCl_2$ and the mixture heated at 55°C for 5 min to denature cohesive termini. Upon cooling to room temperature, 2μl 10mM ATP, 30mM DTT and 1μl T4 DNA ligase (1U) were added and the reaction incubated overnight at 16°C. TE buffer was added to a final volume of 50μl and DNA precipitated with 150μl ethanol. Pelleted DNA was dissolved in 10μl TE and 1μl of the solution used to transform *E. coli* XL1-Blue (Bullock *et al*, 1987) by electroporation. Transformants were selected on LB plates containing ampicillin (50μgml^{-1}), IPTG (0.1mM) and X-gal (80μgml^{-1}). Clones were screened following DNA extraction by restriction mapping and DNA sequence analysis.

The restriction enzymes used to digest pBluescript II SK$^+$ for the cloning of each of these blocks were as

follows:

Block	pBluescript II SK+ digested with:
1	*Kpn*I, *Bam*HI
2	*Kpn*I, *Hind*III
3	*Hind*III, *Not*I
4	*Not*I, *Sac*I
5	*Spe*I, *Sac*I
6	*Kpn*I, *Spe*I
7	*Kpn*I, *Pst*I
8	*Bam*HI, *Pst*I

Construction of pSHELα,ß and γ

Two (pSHELγ) or three (pSHELα,ß) blocks were ligated into pBluescript II SK+ in a single reaction. Each block was excised from the appropriate pBluescript II SK+ -derived plasmid and purified by preparative agarose gel electrophoresis. 25ng (*ca.* 0.15pmol) of each synthetic block (eg. blocks 1-3 in the case of pSHELα) and 150ng (*ca.* 0.075pmol) of the appropriate pBluescript II SK+ fragment were ligated in a total reaction volume of 20μl under conditions similar to those used to assemble the individual blocks. Transformants were screened by restriction analysis. The digestion schemes are illustrated in Figure 2.

Final Assembly of the SHEL gene

The three gene subassemblies pSHELα,ß and γ were excised from their parent plasmids by treatment with the appropriate restriction enzymes (see cloning scheme) and purified by agarose gel electrophoresis. 100ng of pBluescript II SK+ DNA linearised with *Bam*Hl and treated with calf alkaline phosphatase. This and 50ng (*ca.* 0.10pmol) of each subassembly were ligated at 16°C for 1 hour using the DNA Ligation Kit (Amersham International plc) according to the supplied protocol. Transformants were selected on LB-ampicillin plates containing IPTG and X-gal, and analysed by restriction mapping. The two

orientations of the SHEL gene in pBluescript were designated pSHELA and pSHELB.

Expression

The full length SHEL gene was excised from pSHELB with *Bam*HI and purified by gel electrophoresis. 200ng of the purified fragment was ligated with 100ng pGEX-2T linearized with *Bam*HI and treated with calf alkaline phosphatase using the DNA Ligation Kit (Amersham International plc) according to the supplied protocol. Transformants were selected on LB-ampicillin plates and screened by restriction mapping. The SHEL gene cloned into pGEX-2T was designated pSHELC.

Small scale expression of pSHELC was achieved by growing 5ml cultures of *E.coli* DH5α containing pSHELC in LB with 50μg/ml ampicillin and 0.2% glucose at 37°C overnight. 250μl was subinoculated into 5ml 2TY and grown to an A_{600} of approximately 0.8 before being induced with 1mM IPTG. Cultures were grown for a further 3 hours before harvesting. For the analysis of total cell protein 1ml culture was harvested by centrifugation and resuspended in 200μl SDS-PAGE loading buffer. 20μl samples were boiled for 5 minutes before being analysed on an 8% SDS-PAGE gel. For the analysis of soluble and insoluble protein, the bacterial pellet from 3ml culture was resuspended in 500μl lysis buffer (50mM Tris-HCl pH 8, 1mM EDTA, 100mM NaCl) and lysed by the addition of 1mg/ml lysozyme at 4°C for 30 minutes followed by 1% triton X-100 for 20 minutes. After the addition of 0.1 mg/ml DNase samples were sonicated. The samples were centrifuged for 15 minutes in a microfuge and the pellet resuspended in an identical volume of lysis buffer as supernatant. 20μl samples of supernatant and resuspended pellet were boiled for 5 minutes and analysed by 8% SDS-PAGE. (Figure 7). The calculated size of the protein from SDS-PAGE was 86kD which is in close agreement with the predicted size of 90kD. The protein was over 75% soluble under the conditions used. Total amino acid

content of the fusion protein was determined and the results show a high correlation with the predicted values (Figure 8). The total level of expression was determined using SDS-PAGE and scanning densitometry and was found to be in excess of 100 mg/l.

After purification of GST away from SHEL a yield of up to 70 mg/l could theoretically be obtained.

Even allowing for losses during purification this is a highly significant improvement over 4 mg/l obtained with cDNA clones (Indik *et al* 1990). Optimising codon preference has therefore increased the potential yield of tropoelastin fifteenfold.

Alternatively, the SHEL gene was excised from pSHELB with both *Nco*I and *Bam*HI and purified as above. 100ng of the purified fragment was ligated to 50ng pET3d, previously digested with *Nco*I and *Bam*HI, using the Amersham DNA Ligation Kit to give pSHELF. pSHELF was used to transform *E.coli* HMS174. After confirmation, pSHELF was extracted from HMS174 and used to transform BL21. In both cases, transformants were selected on LB-ampicillin plates and screened by restriction mapping.

For pSHELF expression, 5ml LB containing $50\mu gml^{-1}$ ampicillin was inoculated with a single colony of *E.coli* BL21 (DE3) containing pSHELF and incubated overnight at 37°C with shaking. 0.25ml of this culture was used to inoculate 5ml fresh LB containing $50\mu gml^{-1}$ ampicillin and grown to early log phase (A_{600}=0.8 approx). IPTG was added to a final concentration of 0.4mM and growth continued for a further 3h. Total cellular protein was analysed as for pSHELC. Cell lysates were prepared by resuspension of the cell pellet in 9 volumes lysis buffer and incubation at 4°C for 30min with $1mgml^{-1}$ lysozyme. PMSF was added to 0.5mM before the mixture was twice frozen in liquid nitrogen and thawed at 37°C. DNase was added to a concentraton of $0.1mgml^{-1}$ with 10mM $MgCl_2$ and incubated for 20min at room temperature or until the solution was no longer viscous. Insoluble material was removed by centrifugation at 20 000rpm for 25min.

The soluble cell lysate from 125ml culture was extracted by use of a modified version of a technique previously described for tropoelastin isolation (Sandberg *et al.*, 1971). 1.5 volumes of n-propanol was added to the lysate in five aliquots over 2 hours followed by 2.5 volumes of n-butanol. All additions were performed at 4°C with constant stirring and the mixture was allowed to extract overnight. The precipitated protein was removed by centrifugation for 15min at 10 000rpm. The soluble alcohol fraction was frozen and dried *via* a vacuum pump coupled to a liquid nitrogen trap. The residue was dissolved in 3.5ml 25mM HEPES pH 8.0 and dialyzed against 1 l of the same buffer for 2 hours, changed to fresh buffer and dialyzed overnight. The butanol precipitated protein was dissolved in an identical volume SDS-PAGE loading buffer and both fractions were analyzed by SDS-PAGE.

The butanol-extracted protein containing SHEL was further purified by size fractionation using a Superose 12 column and FPLC (Pharmacia). Protein was eluted using 25mM HEPES, pH 8.0. at a flow rate of 0.5 $mlmin^{-1}$.

Protein concentration was estimated using a Bradford assay (Ausubel *et al.*, 1989).

Scanning densitometry of gels was performed on a Molecular Dynamics Personal Densitometer and analyzed using ImageQuant software.

From SDS-PAGE the directly-expressed SHEL was calculated as being 64kDa (Figure 10) which is as predicted. Total amino acid content was determined and was found to be in close agreement with predictions further confirming the nature of the overexpressed protein. The analysis (Figure 11) performed omits lysine residues.

Scanning densitometry of gels was used to estimate the relative level of overexpression. SHEL was expressed at a level of approximately 17% total cell protein in the range 20-200kDa. This represents a substantial level of overexpression and confirms the value of codon

manipulation for high level expression.

As a result of the high levels of expression large quantities of tropoelastin were obtained which can be used for further studies. The directly expressed SHEL protein appeared stable and the rapid degradation seen previously with cDNA expression (Indik *et al.*, 1990) was not observed. Therefore, the purification of the free polypeptide was pursued in preference to fusion protein. A technique utilizing tropoelastin's high solubility in short-chained alcohols has been used previously in the extraction and purification of tropoelastin from tissues (Sandberg *et al.*, 1971). This method was modified for use with soluble cell lysates and found to be very effective. SHEL was selectively extracted into the alcohols while the majority of contaminating protein was precipitated and removed (Fig. 10). The yield of SHEL after this step was high (greater than 90%) despite some loss (less than 10%) by precipitation. The resulting SHEL was of high purity as judged by SDS-PAGE after Coomassie staining (estimated by eye to be of the order greater than 80%). A gel filtration step was used to remove the contaminating protein after which the SHEL was of sufficient purity for further characterization.

Cross-linking of tropoelastin

Tropoelastin obtained from PSHELF (0.3 mg/ml) was chemically cross-linked using 1mM dithiobis (succinimidylpropionate) at 37°C to generate an insoluble material with elastin-like properties. Cross-linking was demonstrated by boiling in the presence of sodium dodecyl sulphate (SDS) followed by SDS-polyacrylamide gel electrophoresis. Cross-linked material did not enter the gel under conditions designed to allow entry of uncross-linked material.

Industrial Applications

Cosmetic Applications

Recombinant tropoelastin is similar or identical to

material found in skin and other tissues and involves no animal death in order to make it. It adds to our own skin's supply of tropoelastin. Recombinant tropoelastins can be used in humans or animals.

Additionally, methods such as liposome technology may be considered to deliver substances deep within the skin.

Another significant area of use for tropoelastin is in minimising scar formation. The availability of large amounts of recombinant tropoelastin means that it should be possible to test whether the scarring obtained from severe cuts and burns can be minimised by regular application of tropoelastin to the affected area. Increased skin elasticity will counter the rigid effects of collagen buildup associated with scar formation, both in human and veterinary applications.

Surgical and Veterinary Applications

The tropoelastins and variants of this invention may be used in the repair and treatment of elastic and non-elastic tissues. They may also be used as food supplements.

REFERENCES

1. Bressan, G.M., Argos, P. and Stanley, K.K. (1987) *Biochemistry* **26**, 1497-1503.

2. Fazio, M.J., Olsen, E.A., Kauh, E.A., Baldwin, C.T., Indik, Z., Ornstein-Goldstein, N., Yeh, H., Rosenbloom, J. and Uitto, J. (1988) *J. Invest. Dermatol.* **91**, 458-464.

3. Indik, Z., Yeh, H., Ornstein-Goldstein, N., Sheppard, P., Anderson, N., Rosenbloom, J.C., Peltonen, L. and Rosenbloom, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5680-5684.

4. Indik, Z., Abrams, W.R., Kucich, U., Gibson, C.W., Mecham, R.P. and Rosenbloom, J. (1990) *Arch. Biochem. Biophys.* **280**, 80-86.

5. Oliver, L., Luvalle, P.A., Davidson, J.M., Rosenbloom, J., Mathew, C.G., Bester, A.J. and Boyd, C.D. (1987) *Collagen Rel. Res.* **7**, 77-89.

6. Alting-Mees, M.A. and Short, J.M. (1989) *Nucl. Acids Res.* **17**, 9494-9494.

7. Bullock, W.O., Fernandez, J.M. and Short, J.M. (1987) *BioTechniques* **5**, 376-379.

8. Gough, J. and Murray, N. (1983) *J.Mol.Biol.* **166**, 1-19.

9. Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) *Nucl. Acids Res.* **16**, 7583-7600.

10. Smith, D.B. and Johnson, K.S. (1988) *Gene* **67**, 31-40.

11. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol.* **185**, 60-89.

12. Lipman and Pearce (1985) Science 227,1435.

13. Heim, R.A., Pierce, R.A., Deak, S.B., Riley, D.J., Boyd, C.D. and Stolle, C.A., (1991) Matrix II 359-366.

14. Raju, K., and Anwar, R.A., (1987) J. Biol. Chem. 262 5755-5762.

15. Yeh, H., Ornstein-Goldstein N., Indik, Z., Sheppard, P., Anderson, N., Rosenbloom, J.C., Cicila, G., Yoon, K. and Rosenbloon, J., (1987) Coll. Relat. Res. 7 235-247.

16. Zhang, S., Zubay, G. and Goldman, E., (1991) Gene 105 61-72.

17. Newgard, C.B., Nakano, K., Hwang, P.K., and Fletterick, R.J., (1986) PNAS (USA) 83: 8132-8136.

18. Murray, E.E., Lotzer, J., Eberle, M. (1989) Nucleic Acids Res. 17 477-498.

19. Sambrook, J., Fritsch E.F., and Maniatis, T., (1989) Molecular cloning: a laboratory manual, second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

20. Yoshifuji S; Tanaka K; and Nitto Y (1987) Chem Pharm Bull 35 2994-3000

21. Urry D.W, Haynes B and Harris RD (1986) Biochem Biophys Res. Comm 141 749-55

22. Rapaka RS; Okamoto K., Long M.M. and Urry D.W. (1983) International Journal of Peptide and Protein Research 21 352 - 363.

23. Bedell-Hogan D., Trackman P., Abrams W. Rosenbloom J and Kagan H (1993) J Biol Chem 268 10345 - 10350.

24. Sandberg L.B., Zeikus R.D. and Coltrain I.M. (1971) Biochem Biophys Acta 236 542-545.

25. Ausubel F.M., Brent R., Kingston R.E., Moore D.D., Seidman J.G., Smith J.A. and Struhl K. (1987) Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience, U.S.A.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: WEISS, ANTHONY S
MARTIN, STEPHEN L
UNIVERSITY, SYDNEY

(ii) TITLE OF INVENTION: SYNTHETIC POLYNUCLEOTIDES

(iii) NUMBER OF SEQUENCES: 27

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: GRIFFITH HACK & CO
(B) STREET: LEVEL 8, 168 WALKER STREET
(C) CITY: NORTH SYDNEY
(D) STATE: NEW SOUTH WALES
(E) COUNTRY: AUSTRALIA
(F) ZIP: 2060

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: AU
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: AU PL6520
(B) FILING DATE: 22-DEC-1992

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: AU PL9661
(B) FILING DATE: 28-JUN-1993

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: KURTS, ANN D
(C) REFERENCE/DOCKET NUMBER: 4828WP:ADK

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 61 2 957 5944
(B) TELEFAX: 61 2 957 6288
(C) TELEX: 26547

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2210 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GATCCATGGG TGGCGTTCCG GGTGCTATCC CGGGTGGCGT TCCGGGTGGT GTATTCTACC 60
CAGGCGCGGG TCTGGGTGCA CTGGGCGGTG GTGCGCTGGG CCCGGGTGGT AAACCGCTGA 120
AACCGGTTCC AGGCGGTCTG GCAGGTGCTG GTCTGGGTGC AGGTCTGGGC GCGTTCCCGG 180
CGGTTACCTT CCCGGGTGCT CTGGTTCCGG GTGGCGTTGC AGACGCAGCT GCTGCGTACA 240
AAGCGGCAAA GGCAGGTGCG GGTCTGGGCG GGGTACCAGG TGTTGGCGGT CTGGGTGTAT 300
CTGCTGGCGC AGTTGTTCCG CAGCCGGGTG CAGGTGTAAA ACCGGGCAAA GTTCCAGGTG 360
TTGGTCTGCC GGGCGTATAC CCGGGTGGTG TTCTGCCGGG CGCGCGTTTC CCAGGTGTTG 420
GTGTACTGCC GGGCGTTCCG ACCGGTGCAG GTGTTAAACC GAAGGCACCA GGTGTAGGCG 480
GCGCGTTCGC GGGTATCCCG GGTGTTGGCC CGTTCGGTGG TCCGCAGCCA GGCGTTCCGC 540
TGGGTTACCC GATCAAAGCG CCGAAGCTTC CAGGTGGCTA CGGTCTGCCG TACACCACCG 600
GTAAACTGCC GTACGGCTAC GGTCCGGGTG GCGTAGCAGG TGCTGCGGGT AAAGCAGGCT 660
ACCCAACCGG TACTGGTGTT GGTCCGCAGG CTGCTGCGGC AGCTGCGGCG AAGGCAGCAG 720
CAAAATTCGG CGCGGGTGCA GCGGGTGTTC TGCCGGGCGT AGGTGGTGCT GGCGTTCCGG 780
GTGTTCCAGG TGCGATCCCG GGCATCGGTG GTATCGCAGG CGTAGGTACT CCGGCGGCCG 840
CTGCGGCTGC GGCAGCTGCG GCGAAAGCAG CTAAATACGG TGCGGCAGCA GGCCTGGTTC 900
CGGGTGGTCC AGGCTTCGGT CCGGGTGTTG TAGGCGTTCC GGGTGCTGGT GTTCCGGGCG 960
TAGGTGTTCC AGGTGCGGGC ATCCCGGTTG TACCGGGTGC AGGTATCCCG GGCGCTGCGG 1020
TTCCAGGTGT TGTATCCCCG GAAGCGGCAG CTAAGGCTGC TGCGAAAGCT GCGAAATACG 1080
GAGCTCGTCC GGGCGTTGGT GTTGGTGGCA TCCCGACCTA CGGTGTAGGT GCAGGCGGTT 1140
TCCCAGGTTT CGGCGTTGGT GTTGGTGGCA TCCCGGGTGT AGCTGGTGTT CCGTCTGTTG 1200
GTGGCGTACC GGGTGTTGGT GGCGTTCCAG GTGTAGGTAT CTCCCCGGAA GCGCAGGCAG 1260
CTGCGGCAGC TAAAGCAGCG AAGTACGGCG TTGGTACTCC GGCGGCAGCA GCTGCTAAAG 1320
CAGCGGCTAA AGCAGCGCAG TTCGGACTAG TTCCGGGCGT AGGTGTTGCG CCAGGTGTTG 1380
GCGTAGCACC GGGTGTTGGT GTTGCTCCGG GCGTAGGTCT GGCACCGGGT GTTGGCGTTG 1440
CACCAGGTGT AGGTGTTGCG CCGGGCGTTG GTGTAGCACC GGGTATCGGT CCGGGTGGCG 1500
TTGCGGCTGC TGCGAAATCT GCTGCGAAGG TTGCTGCGAA AGCGCAGCTG CGTGCAGCAG 1560
CTGGTCTGGG TGCGGGCATC CCAGGTCTGG GTGTAGGTGT TGGTGTTCCG GGCCTGGGTG 1620
TAGGTGCAGG GGTACCGGGC CTGGGTGTTG GTGCAGGCGT TCCGGGTTTC GGTGCTGGCG 1680
CGGACGAAGG TGTACGTCGT TCCCTGTCTC CAGAACTGCG TGAAGGTGAC CCGTCCTCTT 1740
CCCAGCACCT GCCGTCTACC CCGTCCTCTC CACGTGTTCC GGGCGCGCTG GCTGCTGCGA 1800
AAGCGGCGAA ATACGGTGCA GCGGTTCCGG GTGTACTGGG CGGTCTGGGT GCTCTGGGCG 1860
GTGTTGGTAT CCCGGGCGGT GTTGTAGGTG CAGGCCCAGC TGCAGCTGCT GCTGCGGCAA 1920
AGGCAGCGGC GAAAGCAGCT CAGTTCGGTC TGGTTGGTGC AGCAGGTCTG GGCGGTCTGG 1980
GTGTTGGCGG TCTGGGTGTA CCGGGCGTTG GTGGTCTGGG TGGCATCCCG CCGGCGGCGG 2040
CAGCTAAAGC GGCTAAATAC GGTGCAGCAG GTCTGGGTGG CGTTCTGGGT GGTGCTGGTC 2100
AGTTCCCACT GGGCGGTGTA GCGGCACGTC CGGGTTTCGG TCTGTCCCCG ATCTTCCCAG 2160
GCGGTGCATG CCTGGGTAAA GCTTGCGGCC GTAAACGTAA ATAATGATAG 2210

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 733 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Ser Met Gly Gly Val Pro Gly Ala Ile Pro Gly Gly Val Pro Gly Gly
1               5                  10                  15
Val Phe Tyr Pro Gly Ala Gly Leu Gly Ala Leu Gly Gly Gly Ala Leu
            20                  25                  30
Gly Pro Gly Gly Lys Pro Leu Lys Pro Val Pro Gly Gly Leu Ala Gly
        35                  40                  45
Ala Gly Leu Gly Ala Gly Leu Gly Ala Phe Pro Ala Val Thr Phe Pro
    50                  55                  60
Gly Ala Leu Val Pro Gly Gly Val Ala Asp Ala Ala Ala Ala Tyr Lys
65                  70                  75                  80
Ala Ala Lys Ala Gly Ala Gly Leu Gly Gly Val Pro Gly Val Gly Gly
                85                  90                  95
Leu Gly Val Ser Ala Gly Ala Val Val Pro Gln Pro Gly Ala Gly Val
            100                 105                 110
Lys Pro Gly Lys Val Pro Gly Val Gly Leu Pro Gly Val Tyr Pro Gly
        115                 120                 125
Gly Val Leu Pro Gly Ala Arg Phe Pro Gly Val Gly Val Leu Pro Gly
    130                 135                 140
Val Pro Thr Gly Ala Gly Val Lys Pro Lys Ala Pro Gly Val Gly Gly
145                 150                 155                 160
Ala Phe Ala Gly Ile Pro Gly Val Gly Pro Phe Gly Gly Pro Gln Pro
                165                 170                 175
Gly Val Pro Leu Gly Tyr Pro Ile Lys Ala Pro Lys Leu Pro Gly Gly
            180                 185                 190
Tyr Gly Leu Pro Tyr Thr Thr Gly Lys Leu Pro Tyr Gly Tyr Gly Pro
        195                 200                 205
Gly Gly Val Ala Gly Ala Ala Gly Lys Ala Gly Tyr Pro Thr Gly Thr
    210                 215                 220
Gly Val Gly Pro Gln Ala Ala Ala Ala Ala Ala Ala Lys Ala Ala Ala
225                 230                 235                 240
Lys Phe Gly Ala Gly Ala Ala Gly Val Leu Pro Gly Val Gly Gly Ala
                245                 250                 255
Gly Val Pro Gly Val Pro Gly Ala Ile Pro Gly Ile Gly Gly Ile Ala
            260                 265                 270
Gly Val Gly Thr Pro Ala Ala Ala Ala Ala Ala Ala Ala Ala Lys
        275                 280                 285
Ala Ala Lys Tyr Gly Ala Ala Ala Gly Leu Val Pro Gly Gly Pro Gly
    290                 295                 300
```

Phe Gly Pro Gly Val Val Gly Val Pro Gly Ala Gly Val Pro Gly Val
305 310 315 320
Gly Val Pro Gly Ala Gly Ile Pro Val Val Pro Gly Ala Gly Ile Pro
325 330 335
Gly Ala Ala Val Pro Gly Val Val Ser Pro Glu Ala Ala Ala Lys Ala
340 345 350
Ala Ala Lys Ala Ala Lys Tyr Gly Ala Arg Pro Gly Val Gly Val Gly
355 360 365
Gly Ile Pro Thr Tyr Gly Val Gly Ala Gly Gly Phe Pro Gly Phe Gly
370 375 380
Val Gly Val Gly Gly Ile Pro Gly Val Ala Gly Val Pro Ser Val Gly
385 390 395 400
Gly Val Pro Gly Val Gly Gly Val Pro Gly Val Gly Ile Ser Pro Glu
405 410 415
Ala Gln Ala Ala Ala Ala Ala Lys Ala Ala Lys Tyr Gly Val Gly Thr
420 425 430
Pro Ala Ala Ala Ala Ala Lys Ala Ala Ala Lys Ala Ala Gln Phe Gly
435 440 445
Leu Val Pro Gly Val Gly Val Ala Pro Gly Val Gly Val Ala Pro Gly
450 455 460
Val Gly Val Ala Pro Gly Val Gly Leu Ala Pro Gly Val Gly Val Ala
465 470 475 480
Pro Gly Val Gly Val Ala Pro Gly Val Gly Val Ala Pro Gly Ile Gly
485 490 495
Pro Gly Gly Val Ala Ala Ala Ala Lys Ser Ala Ala Lys Val Ala Ala
500 505 510
Lys Ala Gln Leu Arg Ala Ala Ala Gly Leu Gly Ala Gly Ile Pro Gly
515 520 525
Leu Gly Val Gly Val Gly Val Pro Gly Leu Gly Val Gly Ala Gly Val
530 535 540
Pro Gly Leu Gly Val Gly Ala Gly Val Pro Gly Phe Gly Ala Gly Ala
545 550 555 560
Asp Glu Gly Val Arg Arg Ser Leu Ser Pro Glu Leu Arg Glu Gly Asp
565 570 575
Pro Ser Ser Ser Gln His Leu Pro Ser Thr Pro Ser Ser Pro Arg Val
580 585 590
Pro Gly Ala Leu Ala Ala Ala Lys Ala Ala Lys Tyr Gly Ala Ala Val
595 600 605
Pro Gly Val Leu Gly Gly Leu Gly Ala Leu Gly Gly Val Gly Ile Pro
610 615 620
Gly Gly Val Val Gly Ala Gly Pro Ala Ala Ala Ala Ala Ala Ala Lys
625 630 635 640
Ala Ala Ala Lys Ala Ala Gln Phe Gly Leu Val Gly Ala Ala Gly Leu
645 650 655
Gly Gly Leu Gly Val Gly Gly Leu Gly Val Pro Gly Val Gly Gly Leu
660 665 670

Gly Gly Ile Pro Pro Ala Ala Ala Ala Lys Ala Ala Lys Tyr Gly Ala
675 680 685
Ala Gly Leu Gly Gly Val Leu Gly Gly Ala Gly Gln Phe Pro Leu Gly
690 695 700
Gly Val Ala Ala Arg Pro Gly Phe Gly Leu Ser Pro Ile Phe Pro Gly
705 710 715 720
Gly Ala Cys Leu Gly Lys Ala Cys Gly Arg Lys Arg Lys
725 730

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 90 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GATCCATGGG TGGCGTTCCG GGTGCTATCC CGGGTGGCGT TCCGGGTGGT GTATTCTACC 60
CAGGCGCGGG TCTGGGTGCA CTGGGCGGTG 90

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 90 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTGCGCTGGG CCCGGGTGGT AAACCGCTGA AACCGGTTCC AGGCGGTCTG GCAGGTGCTG 60
GTCTGGGTGC AGGTCTGGGC GCGTTCCCGG 90

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGGTTACCTT CCCGGGTGCT CTGGTTCCGG GTGGCGTTGC AGACGCAGCT GCTGCGTACA 60
AAGCGGCAAA GGCAGGTGCG GGTCTGGGCG GGGTAC 96

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CAGGTGTTGG CGGTCTGGGT GTATCTGCTG GCGCAGTTGT TCCGCAGCCG GGTGCAGGTG 60

TAAAACCGGG CAAAGTTCCA GGTGTTGGTC TGCCGGGCG 99

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 90 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TATACCCGGG TGGTGTTCTG CCGGGCGCGC GTTTCCCAGG TGTTGGTGTA CTGCCGGGCG 60

TTCCGACCGG TGCAGGTGTT AAACCGAAGG 90

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CACCAGGTGT AGGCGGCGCG TTCGCGGGTA TCCCGGGTGT TGGCCCGTTC GGTGGTCCGC 60

AGCCAGGCGT TCCGCTGGGT TACCCGATCA AAGCGCCGA 99

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AGCTTCCAGG TGGCTACGGT CTGCCGTACA CCACCGGTAA ACTGCCGTAC GGCTACGGTC 60

CGGGTGGCGT AGCAGGTGCT GCGGGTAA 88

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 90 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCAGGCTAC CCAACCGGTA CTGGTGTTGG TCCGCAGGCT GCTGCGGCAG CTGCGGCGAA 60

GGCAGCAGCA AAATTCGGCG CGGGTGCAGC 90

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 93 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGGTGTTCTG CCGGGCGTAG GTGGTGCTGG CGTTCCGGGT GTTCCAGGTG CGATCCCGGG 60

CATCGGTGGT ATCGCAGGCG TAGGTACTCC GGC 93

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGCCGCTGCG GCTGCGGCAG CTGCGGCGAA AGCAGCTAAA TACGGTGCGG CAGCAGGCCT 60

GGTTCCGGGT GGTCCAGGCT TCGGT 85

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 85 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCGGGTGTTG TAGGCGTTCC GGGTGCTGGT GTTCCGGGCG TAGGTGTTCC AGGTGCGGGC 60

ATCCCGGTTG TACCGGGTGC AGGTA 85

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCCCGGGCGC TGCGGTTCCA GGTGTTGTAT CCCCGGAAGC GGCAGCTAAG GCTGCTGCGA 60

AAGCTGCGAA ATACGGAGCT 80

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGTCCGGGCG TTGGTGTTGG TGGCATCCCG ACCTACGGTG TAGGTGCAGG CGGTTTCCCA 60

GGTTTCGGCG TTGGTGTTGG TGGCATCCCG GG 92

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 90 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TGTAGCTGGT GTTCCGTCTG TTGGTGGCGT ACCGGGTGTT GGTGGCGTTC CAGGTGTAGG 60

TATCTCCCCG GAAGCGCAGG CAGCTGCGGC 90

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGCTAAAGCA GCGAAGTACG GCGTTGGTAC TCCGGCGGCA GCAGCTGCTA AAGCAGCGGC 60

TAAAGCAGCG CAGTTCGGA 79

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 94 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTAGTTCCGG GCGTAGGTGT TGCGCCAGGT GTTGGCGTAG CACCGGGTGT TGGTGTTGCT 60

CCGGGCGTAG GTCTGGCACC GGGTGTTGGC GTTG 94

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 95 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CACCAGGTGT AGGTGTTGCG CCGGGCGTTG GTGTAGCACC GGGTATCGGT CCGGGTGGCG 60

TTGCGGCTGC TGCGAAATCT GCTGCGAAGG TTGCT 95

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 100 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCGAAAGCGC AGCTGCGTGC AGCAGCTGGT CTGGGTGCGG GCATCCCAGG TCTGGGTGTA 60

GGTGTTGGTG TTCCGGGCCT GGGTGTAGGT GCAGGGGTAC 100

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 86 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CGGGCCTGGG TGTTGGTGCA GGCGTTCCGG GTTTCGGTGC TGGCGCGGAC GAAGGTGTAC 60

GTCGTTCCCT GTCTCCAGAA CTGCGT 86

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 93 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GAAGGTGACC CGTCCTCTTC CCAGCACCTG CCGTCTACCC CGTCCTCTCC ACGTGTTCCG 60

GGCGCGCTGG CTGCTGCGAA AGCGGCGAAA TAC 93

(2) INFORMATION FOR SEQ ID NO:23:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
GGTGCAGCGG TTCCGGGTGT ACTGGGCGGT CTGGGTGCTC TGGGCGGTGT TGGTATCCCG 60
GGCGGTGTTG TAGGTGCAGG CCCAGCTGCA 90
(2) INFORMATION FOR SEQ ID NO:24:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 88 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
GCTGCTGCTG CGGCAAAGGC AGCGGCGAAA GCAGCTCAGT TCGGTCTGGT TGGTGCAGCA 60
GGTCTGGGCG GTCTGGGTGT TGGCGGTC 88
(2) INFORMATION FOR SEQ ID NO:25:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
TGGGTGTACC GGGCGTTGGT GGTCTGGGTG GCATCCCGCC GGCGGCGGCA GCTAAAGCGG 60
CTAAATACGG TGCAGCAGGT CTGGGTGGCG TTCTGGGT 98
(2) INFORMATION FOR SEQ ID NO:26:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: NO
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
GGTGCTGGTC AGTTCCCACT GGGCGGTGTA GCGGCACGTC CGGGTTTCGG TCTGTCCCCG 60
ATCTTCCCAG GCGGTGCATG CCTGGGTAA 89
(2) INFORMATION FOR SEQ ID NO:27:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AGCTTGCGGC CGTAAACGTA AATAATGATA G 31

FILE WEISS1.APP CONTAINS SEQUENCE ID NOS. 1 to 27 OF INTERNATIONAL PATENT APPLICATION BY THE UNIVERSITY OF SYDNEY ET AL ENTITLED SYNTHETIC POLYNUCLEOTIDES.

FILE WEISS2.APP CONTAINS SEQUENCE ID NOS. 28 to 54 OF INTERNATIONAL PATENT APPLICATION BY THE UNIVERSITY OF SYDNEY ET AL ENTITLED SYNTHETIC POLYNUCLEOTIDES.

THESE SEQUENCE ID NOS. 28 to 54 APPEAR IN FILE WEISS2.APP AS SEQUENCE ID NOS. 1 to 27 SINCE THIS SEQUENCE LISTING WAS CREATED USING THE PATENTIN PROGRAM WHICH APPARENTLY HAS A LIMIT OF 50 PROJECTS. CONSEQUENTLY, THE SEQUENCE LISTING HAD TO BE CREATED IN TWO PARTS.

SEQUENCE LISTING

(1) GENERAL INFORMATION:
(i) APPLICANT: WEISS, ANTHONY S
MARTIN, STEPHEN L
UNIVERSITY, SYDNEY
(ii) TITLE OF INVENTION: SYNTHETIC POLYNUCLEOTIDES
(iii) NUMBER OF SEQUENCES: 27
(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: GRIFFITH HACK & CO
(B) STREET: LEVEL 8, 168 WALKER STREET
(C) CITY: NORTH SYDNEY
(D) STATE: NEW SOUTH WALES
(E) COUNTRY: AUSTRALIA
(F) ZIP: 2060
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version
#1.25
(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: AU
(B) FILING DATE:
(C) CLASSIFICATION:
(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: AU PL6520
(B) FILING DATE: 22-DEC-1992
(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: AU PL9661
(B) FILING DATE: 28-JUN-1993
(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: KURTS, ANN D
(C) REFERENCE/DOCKET NUMBER: 4828WP:ADK
(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 61 2 957 5944
(B) TELEFAX: 61 2 957 6288
(C) TELEX: AA 26547

(2) INFORMATION FOR SEQ ID NO:1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 92 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCGCACCACC GCCCAGTGCA CCCAGACCCG CGCCTGGGTA GAATACACCA CCCGGAACGC 60
CACCCGGGAT AGCACCCGGA ACGCCACCCA TG 92

(2) INFORMATION FOR SEQ ID NO:2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TAACCGCCGG GAACGCGCCC AGACCTGCAC CCAGACCAGC ACCTGCCAGA CCGCCTGGAA 60
CCGGTTTCAG CGGTTTACCA CCCGGGCCCA 90

(2) INFORMATION FOR SEQ ID NO:3:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCCGCCCAGA CCCGCACCTG CCTTTGCCGC TTTGTACGCA GCAGCTGCGT CTGCAACGCC 60
ACCCGGAACC AGAGCACCCG GGAAGG 86

(2) INFORMATION FOR SEQ ID NO:4:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGCAGACCA ACACCTGGAA CTTTGCCCGG TTTTACACCT GCACCCGGCT GCGGAACAAC 60
TGCGCCAGCA GATACACCCA GACCGCCAAC ACCTGGTAC 99

(2) INFORMATION FOR SEQ ID NO:5:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 99 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TGGTGCCTTC GGTTTAACAC CTGCACCGGT CGGAACGCCC GGCAGTACAC CAACACCTGG 60
GAAACGCGCG CCCGGCAGAA CACCACCCGG GTATACGCC 99

(2) INFORMATION FOR SEQ ID NO:6:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 98 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCTTCGGCG CTTTGATCGG GTAACCCAGC GGAACGCCTG GCTGCGGACC ACCGAACGGG 60
CCAACACCCG GGATACCCGC GAACGCGCCG CCTACACC 98

(2) INFORMATION FOR SEQ ID NO:7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CCTGCTTTAC CCGCAGCACC TGCTACGCCA CCCGGACCGT AGCCGTACGG CAGTTTACCG 60
GTGGTGTACG GCAGACCGTA GCCACCTGGA 90

(2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACACCCGCTG CACCCGCGCC GAATTTTGCT GCTGCCTTCG CCGCAGCTGC CGCAGCAGCC 60
TGCGGACCAA CACCAGTACC GGTTGGGTAG 90

(2) INFORMATION FOR SEQ ID NO:9:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 91 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGCCGCCGGA GTACCTACGC CTGCGATACC ACCGATGCCC GGGATCGCAC CTGGAACACC 60
CGGAACGCCA GCACCACCTA CGCCCGGCAG A 91

(2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCCTGGACCA CCCGGAACCA GGCCTGCTGC CGCACCGTAT TTAGCTGCTT TCGCCGCAGC 60
TGCCGCAGCC GCAGC 75

(2) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 85 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CACCCGGTAC AACCGGGATG CCCGCACCTG GAACACCTAC GCCCGGAACA CCAGCACCCG 60
GAACGCCTAC AACACCCGGA CCGAA 85

(2) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCGTATTTCG CAGCTTTCGC AGCAGCCTTA GCTGCCGCTT CCGGGGATAC AACACCTGGA 60
ACCGCAGCGC CCGGGATACC TG 82

(2) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
ATGCCACCAA CACCAACGCC GAAACCTGGG AAACCGCCTG CACCTACACC GTAGGTCGGG 60
ATGCCACCAA CACCAACGCC CGGACGAGCT 90
(2) INFORMATION FOR SEQ ID NO:14:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
GCTGCCTGCG CTTCCGGGGA GATACCTACA CCTGGAACGC CACCAACACC CGGTACGCCA 60
CCAACAGACG GAACACCAGC TACACCCGGG 90
(2) INFORMATION FOR SEQ ID NO:15:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
CTAGTCCGAA CTGCGCTGCT TTAGCCGCTG CTTTAGCAGC TGCTGCCGCC GGAGTACCAA 60
CGCCGTACTT CGCTGCTTTA GCTGCCGCA 89
(2) INFORMATION FOR SEQ ID NO:16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(iii) HYPOTHETICAL: YES
(iv) ANTI-SENSE: YES
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
CTGGTGCAAC GCCAACACCC GGTGCCAGAC CTACGCCCGG AGCAACACCA ACACCCGGTG 60
CTACGCCAAC ACCTGGCGCA ACACCTACGC CCGGAA 96
(2) INFORMATION FOR SEQ ID NO:17:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 95 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTTCGCAGCA ACCTTCGCAG CAGATTTCGC AGCAGCCGCA ACGCCACCCG GACCGATACC 60

CGGTGCTACA CCAACGCCCG GCGCAACACC TACAC 95

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 90 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCCTGCACCT ACACCCAGGC CCGGAACACC AACACCTACA CCCAGACCTG GGATGCCCGC 60

ACCCAGACCA GCTGCTGCAC GCAGCTGCGC 90

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ACCTTCACGC AGTTCTGGAG ACAGGGAACG ACGTACACCT TCGTCCGCGC CAGCACCGAA 60

ACCCGGAACG CCTGCACCAA CACCCAGGCC CGGTAC 96

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 81 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGCCGCTTTC GCAGCAGCCA GCGCGCCCGG AACACGTGGA GAGGACGGGG TAGACGGCAG 60

GTGCTGGGAA GAGGACGGGT C 81

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GCTGGGCCTG CACCTACAAC ACCGCCCGGG ATACCAACAC CGCCCAGAGC ACCCAGACCG 60

CCCAGTACAC CCGGAACCGC TGCACCGTAT TT 92

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 98 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CACCCAGACC GCCAACACCC AGACCGCCCA GACCTGCTGC ACCAACCAGA CCGAACTGAG 60

CTGCTTTCGC CGCTGCCTTT GCCGCAGCAG CAGCTGCA 98

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 86 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AACGCCACCC AGACCTGCTG CACCGTATTT AGCCGCTTTA GCTGCCGCCG CCGGCGGGAT 60

GCCACCCAGA CCACCAACGC CCGGTA 86

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AGCTTTACCC AGGCATGCAC CGCCTGGGAA GATCGGGGAC AGACCGAAAC CCGGACGTGC 60

CGCTACACCG CCCAGTGGGA ACTGACCAGC ACCACCCAG 99

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GATCCTATCA TTATTTACGT TTACGGCCGC A 31

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2210 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GATCCATGGG AGGGGTCCCT GGGGCCATTC CTGGTGGAGT TCCTGGAGGA GTCTTTTATC 60

CAGGGGCTGG TCTCGGAGCC CTTGGAGGAG GAGCGCTGGG GCCTGGAGGC AAACCTCTTA 120

AGCCAGTTCC CGGAGGGCTT GCGGGTGCTG GCCTTGGGGC AGGGCTCGGC GCCTTCCCCG 180

CAGTTACCTT TCCGGGGGCT CTGGTGCCTG GTGGAGTGGC TGACGCTGCT GCAGCCTATA 240

AAGCTGCTAA GGCTGGCGCT GGGCTTGGTG GTGTCCCAGG AGTTGGTGGC TTAGGAGTGT 300

CTGCAGGTGC GGTGGTTCCT CAGCCTGGAG CCGGAGTGAA GCCTGGGAAA GTGCCGGGTG 360

TGGGGCTGCC AGGTGTATAC CCAGGTGGCG TGCTCCCAGG AGCTCGGTTC CCCGGTGTGG 420

GGGTGCTCCC TGGAGTTCCC ACTGGAGCAG GAGTTAAGCC CAAGGCTCCA GGTGTAGGTG 480

GAGCTTTTGC TGGAATCCCA GGAGTTGGAC CCTTTGGGGG ACCGCAACCT GGAGTCCCAC 540

TGGGGTATCC CATCAAGGCC CCCAAGCTGC CTGGTGGCTA TGGACTGCCC TACACCACAG 600

GGAAACTGCC CTATGGCTAT GGGCCCGGAG GAGTGGCTGG TGCAGCGGGC AAGGCTGGTT 660

ACCCAACAGG GACAGGGGTT GGCCCCCAGG CAGCAGCAGC AGCGGCAGCT AAAGCAGCAG 720

CAAAGTTCGG TGCTGGAGCA GCCGGAGTCC TCCCTGGTGT TGGAGGGGCT GGTGTTCCTG 780

GCGTGCCTGG GGCAATTCCT GGAATTGGAG GCATCGCAGG CGTTGGGACT CCAGCTGCAG 840

CTGCAGCTGC AGCAGCAGCC GCTAAGGCAG CCAAGTATGG AGCTGCTGCA GGCTTAGTGC 900

CTGGTGGGCC AGGCTTTGGC CCGGGAGTAG TTGGTGTCCC AGGAGCTGGC GTTCCAGGTG 960

TTGGTGTCCC AGGAGCTGGG ATTCCAGTTG TCCCAGGTGC TGGGATCCCA GGTGCTGCGG 1020

TTCCAGGGGT TGTGTCACCA GAAGCAGCTG CTAAGGCAGC TGCAAAGGCA GCCAAATACG 1080

GGGCCAGGCC CGGAGTCGGA GTTGGAGGCA TTCCTACTTA CGGGGTTGGA GCTGGGGGCT 1140

TTCCCGGCTT TGGTGTCGGA GTCGGAGGTA TCCCTGGAGT CGCAGGTGTC CCTAGTGTCG 1200

GAGGTGTTCC CGGAGTCGGA GGTGTCCCGG GAGTTGGCAT TTCCCCCGAA GCTCAGGCAG 1260

CAGCTGCCGC CAAGGCTGCC AAGTACGGAG TGGGGACCCC AGCAGCTGCA GCTGCTAAAG 1320

CAGCCGCCAA AGCCGCCCAG TTTGGGTTAG TTCCTGGTGT CGGCGTGGCT CCTGGAGTTG 1380

GCGTGGCTCC TGGTGTCGGT GTGGCTCCTG GAGTTGGCTT GGCTCCTGGA GTTGGCGTGG 1440

CTCCTGGAGT TGGTGTGGCT CCTGGCGTTG GCGTGGCTCC CGGCATTGGC CCTGGTGGAG 1500

TTGCAGCTGC AGCAAAATCC GCTGCCAAGG TGGCTGCCAA AGCCCAGCTC CGAGCTGCAG 1560

CTGGGCTTGG TGCTGGCATC CCTGGACTTG GAGTTGGTGT CGGCGTCCCT GGACTTGGAG 1620

TTGGTGCTGG TGTTCCTGGA CTTGGAGTTG GTGCTGGTGT TCCTGGCTTC GGGGCAGGTG 1680

CAGATGAGGG AGTTAGGCGG AGCCTGTCCC CTGAGCTCAG GGAAGGAGAT CCCTCCTCCT 1740

CTCAGCACCT CCCCAGCACC CCCTCATCAC CCAGGGTACC TGGAGCCCTG GCTGCCGCTA 1800

AAGCAGCCAA ATATGGAGCA GCAGTGCCTG GGGTCCTTGG AGGGCTCGGG GCTCTCGGTG 1860

GAGTAGGCAT CCCAGGCGGT GTGGTGGGAG CCGGACCCGC CGCCGCCGCT GCCGCAGCCA 1920

AAGCTGCTGC CAAAGCCGCC CAGTTTGGCC TAGTGGGAGC CGCTGGGCTC GGAGGACTCG 1980

GAGTCGGAGG GCTTGGAGTT CCAGGTGTTG GGGGCCTTGG AGGTATACCT CCAGCTGCAG 2040

CCGCTAAAGC AGCTAAATAC GGTGCTGCTG GCCTTGGAGG TGTCCTAGGG GGTGCCGGGC 2100

AGTTCCCACT TGGAGGAGTG GCAGCAAGAC CTGGCTTCGG ATTGTCTCCC ATTTTCCCAG 2160

GTGGGGCCTG CCTGGGGAAA GCTTGTGGCC GGAAGAGAAA ATGATGATAG 2210

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4045 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTCACTGGCC GTCGTTTTAC AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA 60
TCGCCTTGCA GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG CCCGCACCGA 120
TCGCCCTTCC CAACAGTTGC GCAGCCTGAA TGGCGAATGG CGCCTGATGC GGTATTTTCT 180
CCTTACGCAT CTGTGCGGTA TTTCACACCG CATATGGTGC ACTCTCAGTA CAATCTGCTC 240
TGATGCCGCA TAGTTAAGCC AGCCCCGACA CCCGCCAACA CCCGCTGACG CGCCCTGACG 300
GGCTTGTCTG CTCCCGGCAT CCGCTTACAG ACAAGCTGTG ACCGTCTCCG GGAGCTGCAT 360
GTGTCAGAGG TTTTCACCGT CATCACCGAA ACGCGCGAGA CGAAAGGGCC TCGTGATACG 420
CCTATTTTTA TAGGTTAATG TCATGATAAT AATGGTTTCT TAGACGTCAG GTGGCACTTT 480
TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA 540
TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGTAT 600
GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT GCCTTCCTGT 660
TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG 720
AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTCGCCCCGA 780
AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG 840
TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA ATGACTTGGT 900
TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA GAGAATTATG 960
CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA CAACGATCGG 1020
AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA CTCGCCTTGA 1080
TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC 1140
TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC 1200
CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC TTCTGCGCTC 1260
GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC GTGGGTCTCG 1320
CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG TTATCTACAC 1380
GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC 1440
ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT 1500
AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA ATCTCATGAC 1560
CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA 1620
AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC 1680
ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT 1740
AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTTCTT CTAGTGTAGC CGTAGTTAGG 1800
CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA TCCTGTTACC 1860
AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT 1920
ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC CCAGCTTGGA 1980
GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CATTGAGAAA GCGCCACGCT 2040
TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG 2100

CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG GGTTTCGCCA 2160
CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC TATGGAAAAA 2220
CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG CTCACATGTT 2280
CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA 2340
TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA 2400
GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT GCAGCTGGCA 2460
CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG TGAGTTAGCT 2520
CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT TGTGTGGAAT 2580
TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG CCAAGCTTGG 2640
CTGCAGGTGA TGATTATCAG CCAGCAGAGA TTAAGGAAAA CAGACAGGTT TATTGAGCGC 2700
TTATCTTTCC CTTTATTTTT GCTGCGGTAA GTCGCATAAA AACCATTCTT CATAATTCAA 2760
TCCATTTACT ATGTTATGTT CTGAGGGGAG TGAAAATTCC CCTAATTCGA TGAAGATTCT 2820
TGCTCAATTG TTATCAGCTA TGCGCCGACC AGAACCACTT GCCGATCAGC CAAACGTCTC 2880
TTCAGGCCAC TGACTAGCGA TAACTTTCCC CACAACGGAA CAACTCTCAT TGCATGGGAT 2940
CATTGGGTAC TGTGGGTTTA GTGGTTGTAA AAACACCTGA CCGCTATCCC TGATCAGTTT 3000
CTTGAAGGTA AACTCATCAC CCCCAAGTCT GGCTATGCAG AAATCACCTG GCTCAACAGC 3060
CTGCTCAGGG TCAACGAGAA TTAACATTCC GTCAGGAAAG CTTGGCTTGG AGCCTGTTGG 3120
TGCGGTCATG GAATTACCTT CAACCTCAAG CCAGAATGCA GAATCACTGG CTTTTTTGGT 3180
TGTGCTTACC CATCTCTCCG CATCACCTTT GGTAAAGGTT CTAAGCTTAG GTGAGAACAT 3240
CCCTGCCTGA ACATGAGAAA AAACAGGGTA CTCATACTCA CTTCTAAGTG ACGGCTGCAT 3300
ACTAACCGCT TCATACATCT CGTAGATTTC TCTGGCGATT GAAGGGCTAA ATTCTTCAAC 3360
GCTAACTTTG AGAATTTTTG CAAGCAATGC GGCGTTATAA GCATTTAATG CATTGATGCC 3420
ATTAAATAAA GCACCAACGC CTGACTGCCC CATCCCCATC TTGTCTGCGA CAGATTCCTG 3480
GGATAAGCCA AGTTCATTTT TCTTTTTTTC ATAAATTGCT TTAAGGCGAC GTGCGTCCTC 3540
AAGCTGCTCT TGTGTTAATG GTTTCTTTTT TGTGCTCATA CGTTAAATCT ATCACCGCAA 3600
GGGATAAATA TCTAACACCG TGCGTGTTGA CTATTTTACC TCTGGCGGTG ATAATGGTTG 3660
CATGTACTAA GGAGGTTGTA TGGAACAACG CATAACCCTG AAAGATTATG CAATGCGCTT 3720
TGGGCAAACC AAGACAGCTA AAGATCTCTC ACCTACCAAA CAATGCCCCC CTGCAAAAAA 3780
TAAATTCATA TAAAAAACAT ACAGATAACC ATCTGCGGTG ATAAATTATC TCTGGCGGTG 3840
TTGACATAAA TACCACTGGC GGTGATACTG AGCACATCAG CAGGACGCAC TGACCACCAT 3900
GAAGGTGACG CTCTTAAAAA TTAAGCCCTG AAGAAGGGCA GCATTCAAAG CAGAAGGCTT 3960
TGGGGTGTGT GATACGAAAC GAAGCATTGG GATCCTAAGG AGGTTTAAGA TCCATGGGTT 4020
TAAACCTCCT TAGGATCCCC GGGAA 4045

CLAIMS

1. A synthetic polynucleotide encoding the amino acid sequence of a tropoelastin or a variant of the tropoelastin wherein:

all or some of the codons which hamper expression in the expression system in which the polynucleotide is to be expressed are replaced with codons more favourable for expression in the expression system.

2. A synthetic polynucleotide according to claim 1 wherein at least 50% of codons for any particular amino acid are selected to reflect preferred codon usage in the host of choice.

3. A synthetic polynucleotide according to claim 1 or claim 2 wherein the synthetic polynucleotide excludes all or part of the 5' and/or 3' untranslated regions of the tropoelastin gene corresponding to the synthetic polynucleotide.

4. A synthetic polynucleotide according to claim 1 or claim 2 wherein the synthetic polynucleotide excludes all or part of the tropoelastin signal peptide encoding sequence of the corresponding tropoelastin gene.

5. A synthetic polynucleotide according to claim 1 or claim 2 wherein the synthetic polynucleotide is prepared from assembled oligonucleotides incorporating restrition sites to facilitate assembly of the polynucleotide.

6. A synthetic polynucleotide according to claim 1 or claim 2 wherein the expression system is an E. coli expression system or a yeast, or other bacterial expression system or an insect or other eukaryotic cell expression system or a whole organism.

7. A synthetic polynucleotide according to claim 6 wherein the expression system is E. coli and at least 50% of the base changes indicated in Figure 6 have been made.

8. A synthetic polynucleotide according to claim 1 or claim 2 comprising the sequence depicted in Figure

3(1) to 3(5) (SEQ ID NO: 1).

9. A synthetic polynucleotide according to claim 1 or claim 2 fused to a polynucleotide sequence compatible with the host for the expression system.

10. A synthetic polynucleotide according to claim 9 where the compatible sequence is at the 5' end of the polynucleotide molecule.

11. A synthetic polynucleotide according to claim 10 wherein the compatible polynucleotide encodes all or part of glutathione-S-transferase.

12. A recombinant DNA molecule comprising a synthetic polynucleotide according to claim 1 or claim 2 and vector DNA.

13. A recombinant DNA molecule according to claim 12 wherein the vector is selected from the group consisting of pBR322, pBluescript II SK^+, pGEX-2T, pTrc99A, pET3d and derivatives of these vectors.

14. A plasmid selected from the group consisting of pSHELA, pSHELB, pSHELC and pSHELF.

15. A host transformed with a recombinant DNA molecule according to claim 12 or claim 13 or a plasmid according to claim 14.

16. A host according to claim 15 which host is a bacterium, a yeast, an insect cell or other eukaryotic cell, or a whole organism.

17. A host according to claim 16 which is E. coli strain NM522 or XL1-Blue.

18. An expression product of a host according to claim 15, which expression product comprises a tropoelastin or tropoelastin variant.

19. An expression product according to claim 18 which is SHEL or GST-SHEL.

20. A cross-linked expression product according to claim 18.

21. A cross-linked expression product according to claim 20 which is chemically cross-linked.

22. A cross-linked expression product according to claim 20 which is enzymatically cross-linked.

23. A cross-linked expression product according to claim 20 which is cross-linked by gamma irradiation.

24. A composition comprising an expression product according to claim 18 or a cross-linked expression product according to claim 20 together with a pharmaceutically or veterinarally acceptable carrier.

25. A carrier for delivery of an active agent comprising a coacervate of an expression product according to claim 18.

26. A process for the preparation of an expression product according to claim 18 comprising:

providing a transformed host according to claim 16; culturing it under conditions suitable for expression of the expression product; and collecting the expression product.

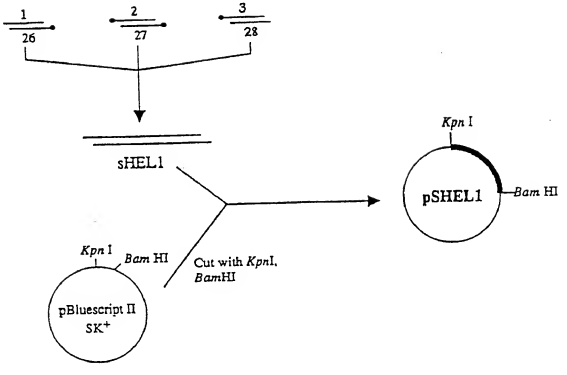
1
26
2
27
3
28
sHEL1
Kpn I
pSHEL1
Bam HI
Kpn I
Bam HI
pBluescript II
SK+
Cut with KpnI,
BamHI

FIG. 1

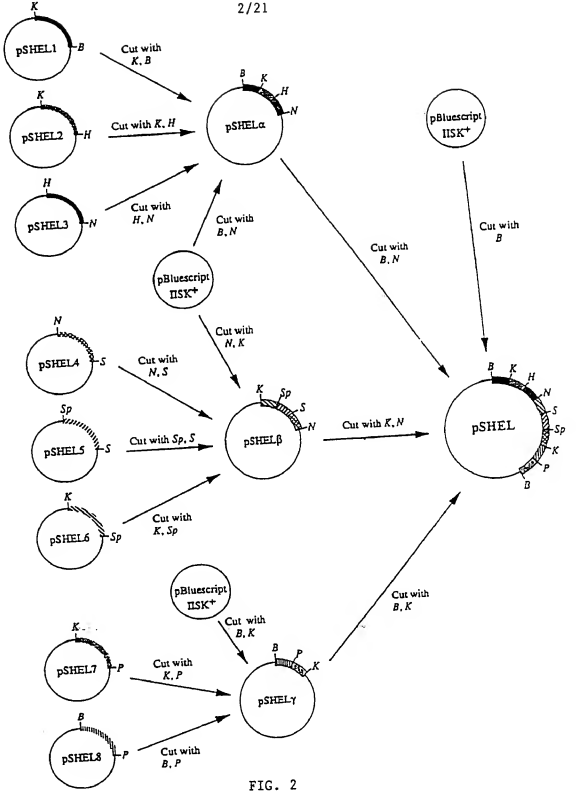
pSHEL1
pSHEL2
pSHEL3
pSHELα
pBluescript IISK+
pSHEL4
pSHEL5
pSHEL6
pSHELβ
pSHEL
pSHEL7
pSHEL8
pSHELγ
Cut with K, B
Cut with K, H
Cut with H, N
Cut with B, N
Cut with N, K
Cut with N, S
Cut with Sp, S
Cut with K, Sp
Cut with K, N
Cut with B
Cut with B, K
Cut with K, P
Cut with B, P

FIG. 2

```
1   GATCCATGGGTGGCGTTCCGGGTGCTATCCCGGGTGGCGTTCCGGGTGGTGTATTCTACC   60
        GTACCCACCGCAAGGCCCACGATAGGGCCCACCGCAAGGGCCCACCACATAAGATGG

        S  M  G  G  V  P  G  A  I  P  G  G  V  P  G  G  V  F  Y  P

61  CAGGCGCGGGTCTGGGTGCACTGGGCGGTGGTGCGCTGGGCCCGGGTGGTAAACCGCTGA  120
    GTCCGCGCCCAGACCCACGTGACCCGCCACCACGCGACCCGGGCCCACCATTTGGCGACT

        G  A  G  L  G  A  L  G  G  G  A  L  G  P  G  G  K  P  L  K

121 AACCGGTTCCAGGCGGTCTGGCAGGTGCTGGTCTGGGTGCAGGTCTGGGCGCGTTCCCGG  180
    TTGGCCAAGGTCCGCCAGACCGTCCACGACCAGACCCACGTCCAGACCCGCGCAAGGGCC

        P  V  P  G  G  L  A  G  A  G  L  G  A  G  L  G  A  F  P  A

181 CGGTTACCTTCCCGGGTGCTCTGGTTCCGGGTGGCGTTGCAGACGGCAGCTGCTGCGTACA  240
    GCCAATGGAAGGGCCCACGAGACCAAGGCCCACCGCAACGTCTGCGTCGACGACGCATGT

        V  T  F  P  G  A  L  V  P  G  G  V  A  D  A  A  A  A  Y  K

241 AAGCGGCAAAGGCAGGTGCGGGTCTGGGCGGGGTACCAGGTGTTGGCGGTCTGGGTGTAT  300
    TTCGCCGTTTCCGTCCACGCCCAGACCCGCCCCATGGTCCACAACCGCCAGACCCACATA

        A  A  K  A  G  A  G  L  G  G  V  P  G  V  G  G  L  G  V  S

301 CTGCTGGCGCAGTTGTTCCGCAGCCGGGTGCAGGTGTAAAACCGGGCAAAGTTCCAGGTG  360
    GACGACCGCGTCAACAAGGCGTCGGCCCACGTCCACATTTTGGCCCGTTTCAAGGTCCAC

        A  G  A  V  V  P  Q  P  G  A  G  V  K  P  G  K  V  P  G  V

361 TTGGTCTGCCGGGCGTATACCCGGGTGGTGTTCTGCCGGGCGCGCGTTTCCCAGGTGTTG  420
    AACCAGACGGCCCGCATATGGGCCCACCACAAGACGGCCCGCGCGCAAAGGGTCCACAAC

        G  L  P  G  V  Y  P  G  G  V  L  P  G  A  R  F  P  G  V  G
```

FIG. 3(1)

```
421  GTGTACTGCCGGGCGTTCCGACCGGTGCAGGTGTTAAACCGAAGGCACCAGGTGTAGGCG  480
     CACATGACGGCCCGCAAGGCTGGCCACGTCCACAATTTGGCTTCCGTGGTCCACATCCGC

       V  L  P  G  V  P  T  G  A  G  V  K  P  K  A  P  G  V  G  G

481  GCGCGTTCGCGGGTATCCCGGGTGTTGGCCCGTTCGGTGGTCCGCAGCCAGGCGTTCCGC  540
     CGCGCAAGCGCCCATAGGGCCCACAACCGGGCAAGCCACCAGGCGTCGGTCCGCAAGGCG

       A  F  A  G  I  P  G  V  G  P  F  G  G  P  Q  P  G  V  P  L

541  TGGGTTACCCGATCAAAGCGCCGAAGCTTCCAGGTGGCTACGGTCTGCCGTACACCACCG  600
     ACCCAATGGGCTAGTTTCGCGGGCTTCGAAGGTCCACCGATGCCAGACGGCATGTGGTGGC

       G  Y  P  I  K  A  P  K  L  P  G  G  Y  G  L  P  Y  T  T  G

601  GTAAACTGCCGTACGGCTACGGTCCGGGTGGCGTAGCAGGTGCTGCGGGTAAAGCAGGCT  660
     CATTTGACGGCATGCCGATGCCAGGCCCACCGCATCGTCCACGACGCCCATTTCGTCCGA

       K  L  P  Y  G  Y  G  P  G  G  V  A  G  A  A  G  K  A  G  Y

661  ACCCAACCGGTACTGGTGTTGGTCCGCAGGCTGCTGCGGCAGCTGCGGCGAAGGCAGCAG  720
     TGGGTTGGCCATGACCACAACCAGGCGTCCGACGACGCCGTCGACGCCGCTTCCGTCGTC

       P  T  G  T  G  V  G  P  Q  A  A  A  A  A  A  A  K  A  A  A

721  CAAAATTCGGCGCGGGTGCAGCGGGTGTTCTGCCGGGCGTAGGTGGTGCTGGCGTTCCGG  780
     GTTTTAAGCCGCGCCCACGTCGCCCACAAGACGGGCCCGCATCCACCACGACCGCAAGGCC

       K  F  G  A  G  A  A  G  V  L  P  G  V  G  G  A  G  V  P  G

781  GTGTTCCAGGTGCGATCCCGGGCATCGGTGGTATCGCAGGCGTAGGTACTCCGGCGGCCG  840
     CACAAGGTCCACGCTAGGGCCCGTAGCCACCATAGCGTCCGCATCCATGAGGCCGCCGGC

       V  P  G  A  I  P  G  I  G  G  I  A  G  V  G  T  P  A  A  A

841  CTGCGGCTGCGGCAGCTGCGGCGAAAGCAGCTAAATACGGTGCGGCAGCAGGCCTGGTTC  900
     GACGCCGACGCCGTCGACGCCGCTTTCGTCGATTTATGCCACGCCGTCGTCCGGACCAAG

       A  A  A  A  A  A  A  K  A  A  K  Y  G  A  A  A  G  L  V  P
```

FIG. 3(2)

```
901  CGGGTGGTCCAGGCTTCGGTCCGGGTGTTGTAGGCGTTCCGGGTGCTGGTGTTCCGGGCG   960
     GCCCACCAGGTCCGAAGCCAGGCCCACAACATCCGCAAGGCCCACGACCACAAGGCCCGC

       G  G  P  G  F  G  P  G  V  V  G  V  P  G  A  G  V  P  G  V

961  TAGGTGTTCCAGGTGCGGGCATCCCGGTTGTACCGGGTGCAGGTATCCCGGGCGCTGCGG   1020
     ATCCACAAGGTCCACGCCCGTAGGGCCAACATGGCCCACGTCCATAGGGCCCGCGACGCC

       G  V  P  G  A  G  I  P  V  V  P  G  A  G  I  P  G  A  A  V

1021 TTCCAGGTGTTGTATCCCCGGAAGCGGCAGCTAAGGCTGCTGCGAAAGCTGCGAAATACG   1080
     AAGGTCCACAACATAGGGGCCTTCGCCGTCGATTCCGACGACGCTTTCGACGCTTTATGC

       P  G  V  V  S  P  E  A  A  A  K  A  A  A  K  A  A  K  Y  G

1081 GAGCTCGTCCGGGGCGTTGGTGTTGGTGGCATCCCGACCTACGGTGTAGGTGCAGGCGGTT   1140
     CTCGAGCAGGCCCGCAACCACAACCACCGTAGGGCTGGATGCCACATCCACGTCCGCCAA

       A  R  P  G  V  G  V  G  G  I  P  T  Y  G  V  G  A  G  G  F

1141 TCCCAGGTTTCGGCGTTGGTGTTGGTGGCATCCCGGGTGTAGCTGGTGTTCCGTCTGTTG   1200
     AGGGTCCAAAGCCGCAACCACAACCACCGTAGGGCCCACATCGACCACAAGGCAGACAAC

       P  G  F  G  V  G  V  G  G  I  P  G  V  A  G  V  P  S  V  G

1201 GTGGCGTACCGGGTGTTGGTGGCGTTCCAGGTGTAGGTATCTCCCCGGAAGCGCAGGCAG   1260
     CACCGCATGGCCCACAACCACCGCAAGGTCCACATCCATAGAGGGGCCTTCGCGTCCGTC

       G  V  P  G  V  G  G  V  P  G  V  G  I  S  P  E  A  Q  A  A

1261 CTGCGGCAGCTAAAGCAGCGAAGTACGGCGTTGGTACTCCGGCGGCAGCAGCTGCTAAAG   1320
     GACGCCGTCGATTTCGTCGCTTCATGCCGCAACCATGAGGCCGCCGTCGTCGACGATTTC

       A  A  A  K  A  A  K  Y  G  V  G  T  P  A  A  A  A  A  K  A

1321 CAGCGGCTAAAGCAGCGCAGTTCGGACTAGTTCCGGGCGTAGGTGTTGCGCCAGGTGTTG   1380
     GTCGCCGATTTCGTCGCGTCAAGCCTGATCAAGGCCCGCATCCACAACGCGGTCCACAAC

       A  A  K  A  A  Q  F  G  L  V  P  G  V  G  V  A  P  G  V  G
```

FIG. 3(3)

```
1381  GCGTAGCACCGGGTGTTGGTGTTGCTCCGGGCGTAGGTCTGGCACCGGGTGTTGGCGTTG   1440
      CGCATCGTGGCCCACAACCACAACGAGGCCCGCATCCAGACCGTGGCCCACAACCGCAAC

          V  A  P  G  V  G  V  A  P  G  V  G  L  A  P  G  V  G  V  A

1441  CACCAGGTGTAGGTGTTGCGCCGGGCGTTGGTGTAGCACCGGGTATCGGTCCGGGTGGCG   1500
      GTGGTCCACATCCACAACGCGGCCCGCAACCACATCGTGGCCCATAGCCAGGCCCACCGC

          P  G  V  G  V  A  P  G  V  G  V  A  P  G  I  G  P  G  G  V

1501  TTGCGGCTGCTGCGAAATCTGCTGCGAAGGTTGCTGCGAAAGCGCAGCTGCGTGCAGCAG   1560
      AACGCCGACGACGCTTTAGACGACGCTTCCAACGACGCTTTCGCGTCGACGCACGTCGTC

          A  A  A  A  K  S  A  A  K  V  A  A  K  A  Q  L  R  A  A  A

1561  CTGGTCTGGGTGCGGGCATCCCAGGTCTGGGTGTAGGTGTTGGTGTTCCGGGCCTGGGTG   1620
      GACCAGACCCACGCCCGTAGGGTCCAGACCCACATCCACAACCACAAGGCCCGGACCCAC

          G  L  G  A  G  I  P  G  L  G  V  G  V  G  V  P  G  L  G  V

1621  TAGGTGCAGGGGTACCGGGCCTGGGTGTTGGTGCAGGCGTTCCGGGTTTCGGTGCTGGCG   1680
      ATCCACGTCCCCATGGCCCGGACCCACAACCACGTCCGCAAGGCCCAAAGCCACGACCGC

          G  A  G  V  P  G  L  G  V  G  A  G  V  P  G  F  G  A  G  A

1681  CGGACGAAGGTGTACGTCGTTCCCTGTCTCCAGAACTGCGTGAAGGTGACCCGTCCTCTT   1740
      GCCTGCTTCCACATGCAGCAAGGGACAGAGGTCTTGACGCACTTCCACTGGGCAGGAGAA

          D  E  G  V  R  R  S  L  S  P  E  L  R  E  G  D  P  S  S  S

1741  CCCAGCACCTGCCGTCTACCCCGTCCTCTCCACGTGTTCCGGGCGCGCTGGCTGCTGCGA   1800
      GGGTCGTGGACGGCAGATGGGGCAGGAGAGGTGCACAAGGCCCGCGCGACCGACGACGCT

          Q  H  L  P  S  T  P  S  S  P  R  V  P  G  A  L  A  A  A  K

1801  AAGCGGCGAAATACGGTGCAGCGGTTCCGGGTGTACTGGGCGGTCTGGGTGCTCTGGGCG   1860
      TTCGCCGCTTTATGCCACGTCGCCAAGGCCCACATGACCCGCCAGACCCACGAGACCCGC

          A  A  K  Y  G  A  A  V  P  G  V  L  G  G  L  G  A  L  G  G
```

FIG. 3(4)

```
1861  GTGTTGGTATCCCGGGCGGTGTTGTAGGTGCAGGCCCAGCTGCAGCTGCTGCTGCGGCAA  1920
      CACAACCATAGGGCCCGCCACAACATCCACGTCCGGGTCGACGTCGACGACGACGCCGTT

        V  G  I  P  G  G  V  V  G  A  G  P  A  A  A  A  A  A  A  K

1921  AGGCAGCGGCGAAAGCAGCTCAGTTCGGTCTGGTTGGTGCAGCAGGTCTGGGCGGTCTGG  1980
      TCCGTCGCCGCTTTCGTCGAGTCAAGCCAGACCAACCACGTCGTCCAGACCCGCCAGACC

        A  A  A  K  A  A  Q  F  G  L  V  G  A  A  G  L  G  G  L  G

1981  GTGTTGGCGGTCTGGGTGTACCGGGCGTTGGTGGTCTGGGTGGCATCCCGCCGGCGGCGG  2040
      CACAACCGCCAGACCCACATGGCCCGCAACCACCAGACCCACCGTAGGGCGGCCGCCGCC

        V  G  G  L  G  V  P  G  V  G  G  L  G  G  I  P  P  A  A  A

2041  CAGCTAAAGCGGCTAAATACGGTGCAGCAGGTCTGGGTGGCGTTCTGGGTGGTGCTGGTC  2100
      GTCGATTTCGCCGATTTATGCCACGTCGTCCAGACCCACCGCAAGACCCACCACGACCAG

        A  K  A  A  K  Y  G  A  A  G  L  G  G  V  L  G  G  A  G  Q

2101  AGTTCCCACTGGGCGGTGTAGCGGCACGTCCGGGTTTCGGTCTGTCCCCGATCTTCCCAG  2160
      TCAAGGGTGACCCGCCACATCGCCGTGCAGGCCCAAAGCCAGACAGGGGCTAGAAGGGTC

        F  P  L  G  G  V  A  A  R  P  G  F  G  L  S  P  I  F  P  G

2161  GCGGTGCATGCCTGGGTAAAGCTTGCGGCCGTAAACGTAAATAATGATAG              2210
      CGCCACGTACGGACCCATTTCGAACGCCGGCATTTGCATTTATTACTATCCTAG
        G  A  C  L  G  K  A  C  G  R  K  R  K  *  *  *
```

FIG. 3(5)

No.	SEQUENCE
1	GATCCATGGGTGGCGTTCCGGGTGCTATCCCGGGTGGCGTTCCGGGTGGTGTATTCTACCCAGGCGCGGGTCTGGGTGCACTGGGCGGTG
2	GTGCGCTGGGCCCGGGTGGTAAACCGCTGAAACCGGTTCCAGGCGGTCTGGCAGGTGCTGGTCTGGGTGCAGGTCTGGGCGCGTTCCCGG
3	CGGTTACCTTCCCGGGTGCTCTGGTTCCGGGTGGCGTTGCAGACGCAGCTGCTGCGTACAAAGCGGCAAAGGCAGGTGCGGGTCTGGGCGGGGTAC
4	CAGGTGTTGGCGGTCTGGGTGTATCTGCTGGCGCAGTTGTTCCGCAGCCGGGTGCAGGTGTAAAACCGGGCAAAGTTCCAGGTGTTGGTCTGCCGGGCG
5	TATACCCGGGTGGTGTTCTGCCGGGCGCGCGTTTCCCAGGTGTTGGTGTACTGCCGGGCGTTCCGACCGGTGCAGGTGTTAAACCGAAGG
6	CACCAGGTGTAGGCGGCGCGTTCGCGGGTATCCCGGGTGTTGGCCCGTTCGGTGGTCCGCAGCCAGGCGTTCCGCTGGGTTACCCGATCAAAGCGCCGA
7	AGCTTCCAGGTGGCTACGGTCTGCCGTACACCACCGGTAAACTGCCGTACGGCTACGGTCCGGGTGGCGTAGCAGGTGCTGCGGGTAA
8	AGCAGGCTACCCAACCGGTACTGGTGTTGGTCCGCAGGCTGCTGCGGCAGCTGCGGCGAAGGCAGCAGCAAAATTCGGCGCGGGTGCAGC
9	GGGTGTTCTGCCGGGCGTAGGTGGTGCTGGCGTTCCGGGTGTTCCAGGTGCGATCCCGGGCATCGGTGGTATCGCAGGCGTAGGTACTCCGGC
10	GGCCGCTGCGGCTGCGGCAGCTGCGGCGAAAGCAGCTAAATACGGTGCGGCAGCAGGCCTGGTTCCGGGTGGTCCAGGCTTCGGT
11	CCGGGTGTTGTAGGCGTTCCGGGTGCTGGTGTTCCGGGCGTAGGTGTTCCAGGTGCGGGCATCCCGGTTGTACCGGGTGCAGGTA

FIG. 4(1)

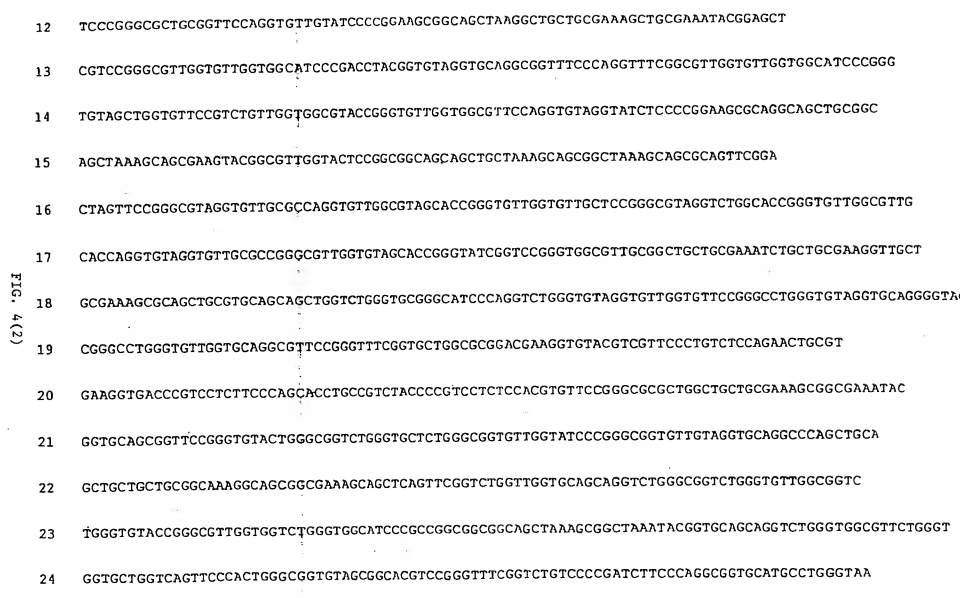
12 TCCCGGGCGCTGCGGTTCCAGGTGTTGTATCCCCGGAAGCGGCAGCTAAGGCTGCTGCGAAAGCTGCGAAATACGGAGCT
13 CGTCCGGGCGTTGGTGTTGGTGGCATCCCGACCTACGGTGTAGGTGCAGGCGGTTTCCCAGGTTTCGGCGTTGGTGTTGGTGGCATCCCGGG
14 TGTAGCTGGTGTTCCGTCTGTTGGTGGCGTACCGGGTGTTGGTGGCGTTCCAGGTGTAGGTATCTCCCCGGAAGCGCAGGCAGCTGCGGC
15 AGCTAAAGCAGCGAAGTACGGCGTTGGTACTCCGGCGGCAGCAGCTGCTAAAGCAGCGGCTAAAGCAGCGCAGTTCGGA
16 CTAGTTCCGGGCGTAGGTGTTGCGCCAGGTGTTGGCGTAGCACCGGGTGTTGGTGTTGCTCCGGGCGTAGGTCTGGCACCGGGTGTTGGCGTTG
17 CACCAGGTGTAGGTGTTGCGCCGGGCGTTGGTGTAGCACCGGGTATCGGTCCGGGTGGCGTTGCGGCTGCTGCGAAATCTGCTGCGAAGGTTGCT
18 GCGAAAGCGCAGCTGCGTGCAGCAGCTGGTCTGGGTGCGGGCATCCCAGGTCTGGGTGTAGGTGTTGGTGTTCCGGGCCTGGGTGTAGGTGCAGGGGTAC
19 CGGGCCTGGGTGTTGGTGCAGGCGTTCCGGGTTTCGGTGCTGGCGCGGACGAAGGTGTACGTCGTTCCCTGTCTCCAGAACTGCGT
20 GAAGGTGACCCGTCCTCTTCCCAGCACCTGCCGTCTACCCCGTCCTCTCCACGTGTTCCGGGCGCGCTGGCTGCTGCGAAAGCGGCGAAATAC
21 GGTGCAGCGGTTCCGGGTGTACTGGGCGGTCTGGGTGCTCTGGGCGGTGTTGGTATCCCGGGCGGTGTTGTAGGTGCAGGCCCAGCTGCA
22 GCTGCTGCTGCGGCAAAGGCAGCGGCGAAAGCAGCTCAGTTCGGTCTGGTTGGTGCAGCAGGTCTGGGCGGTCTGGGTGTTGGCGGTC
23 TGGGTGTACCGGGCGTTGGTGGTCTGGGTGGCATCCCGCCGGCGGCGGCAGCTAAAGCGGCTAAATACGGTGCAGCAGGTCTGGGTGGCGTTCTGGGT
24 GGTGCTGGTCAGTTCCCACTGGGCGGTGTAGCGGCACGTCCGGGTTTCGGTCTGTCCCCGATCTTCCCAGGCGGTGCATGCCTGGGTAA
25 AGCTTGCGGCCGTAAACGTAAATAATGATAG

FIG. 4(2)

No.	SEQUENCE
26	GCGCACCACCGCCCAGTGCACCCAGACCCGCGCCTGGGTAGAATACACCACCCGGAACGCCACCCGGGATAGCACCCGGAACGCCACCCATG
27	TAACCGCCGGGAACGCGCCCAGACCTGCACCCAGACCAGCACCTGCCAGACCGCCTGGAACCGGTTTCAGCGGTTTACCACCCGGGCCCA
28	CCCGCCCAGACCCGCACCTGCCTTTGCCGCTTTGTACGCAGCAGCTGCGTCTGCAACGCCACCCGGAACCAGAGCACCCGGGAAGG
29	CGGCAGACCAACACCTGGAACTTTGCCCGGTTTTACACCTGCACCCGGCTGCGGAACAACTGCGCCAGCAGATACACCCAGACCGCCAACACCTGGTAC
30	TGGTGCCTTCGGTTTAACACCTGCACCGGTCGGAACGCCCGGCAGTACACCAACACCTGGGAAACGCGCGCCCGGCAGAACACCACCCGGGTATACGCC
31	AGCTTCGGCGCTTTGATCGGGTAACCCAGCGGAACGCCTGGCTGCGGACCACCGAACGGGCCAACACCCGGGATACCCGCGAACGCGCCGCCTACACC
32	CCTGCTTTACCCGCAGCACCTGCTACGCCACCCGGACCGTAGCCGTACGGCAGTTTACCGGTGGTGTACGGCAGACCGTAGCCACCTGGA
33	ACACCCGCTGCACCCGCGCCGAATTTTGCTGCTGCCTTCGCCGCAGCTGCCGCAGCAGCCTGCGGACCAACACCAGTACCGGTTGGGTAG
34	GGCCGCCGGAGTACCTACGCCTGCGATACCACCGATGCCCGGGATCGCACCTGGAACACCCGGAACGCCAGCACCACCTACGCCCGGCAGA
35	GCCTGGACCACCCGGAACCAGGCCTGCTGCCGCACCGTATTTAGCTGCTTTCGCCGCAGCTGCCGCAGCCGCAGC
36	CACCCGGTACAACCGGGATGCCCGCACCTGGAACACCTACGCCCGGAACACCAGCACCCGGAACGCCTACAACACCCGGACCGAA
37	CCGTATTTCGCAGCTTTCGCAGCAGCCTTAGCTGCCGCTTCCGGGGATACAACACCTGGAACCGCAGCGCCCGGGATACCTG

FIG. 5(1)

38	ATGCCACCAACACCAACGCCGAAACCTGGGAAACCGCCTGCACCTACACCGTAGGTCGGGATGCCACCAACACCAGCCCGGACGAGCT
39	GCTGCCTGCGCTTCCGGGGAGATACCTACACCTGGAACGCCACCAACACCCGGTACGCCACCAACAGACGGAACACCAGCTACACCCGGG
40	CTAGTCCGAACTGCGCTGCTTTAGCAGCTGCTGCCGCCGGAGTACCAACGCCGTACTTCGCTGCTTTAGCTGCCGCA
41	CTGGTGCAACGCCAACACCCGGTGCCAGACCTACGCCCGGAGCAACACCAACACCCGGTGCTACGCCAACACCTGGCGCAACACCCTACGCCCGGAA
42	TTTCGCAGCAACCTTCGCAGCAGATTTCGCAGCAGCCGCAACGCCCACCCGGACCGATACCCGGTGCTACACCAACGCCCGGCGCAACACCTACAC
43	CCCTGCACCTACACCCAGGCCCGGAACACCAACACCTACACCCAGACCTGGGATGCCCGCCACCCAGACCAGCTGCTGCACGCAGCTGCGC
44	ACCTTCACGCAGTTCTGGAGACAGGGACGACGTACACCTTCGTCCGCGCCAGCACCGAAACCCGGAACGCCTGCACCAACACCCAGGCCCGGTAC
45	CGCCGCTTTCGCAGCAGCCAGCCGCCCGGAACACGTGGAGAGACGGGGTAGACGGCAGGTGCTGGGAAGAGGACGGGTC
46	GCTGGGCCTGCACCTACAACACCGCCCGGGATACCAACACCGCCCAGAGCACCCAGACCGCCACGTACACCCGGAACCGCTGCACCGTATTT
47	CACCCAGACCGCCAACACCCAGACCGCCCAGACCTGCTGCACCAACCAGACCGAACTGAGCTGCTTTCGCCGCTGCCTTTGCCGCAGCAGCGTGCA
48	AACGCCACCCAGACCTGCTGCACCGTATTTAGCCGCTTTAGCTGCCGCCGCCGGGGATGCCACCCACGACCACCAACGCCCGGTA
49	AGCTTTACCCAGGCATGCACCGCCTGGGAAGAGTCGGGGACAGACCGAAACCCGGACGTGCCGCTACACCGCCCAGTGGGAACTGACCAGCACCACCCAG
50	GATCCTATCATATTTACGTTTACGGCCGCA

FIG. 5(2)

```
1   GATCCATGGGTGGCGTTCCGGGTGCTATCCCGGGTGGCGTTCCGGGTGGTGTATTCTACC   60
                A G C T G C T T   A   T A A C T T

61  CAGGCGCGGGTCTGGGTGCACTGGGCGGTGGTGCGCTGGGCCCGGGTGGTAAACCGCTGA  120
      G T   C A C T A A A     G T A C   T T

121 AACCGGTTCCAGGCGGTCTGGCAGGTGCTGGTCTGGGTGCAGGTCTGGGCGCGTTCCCGG  180
     G A   C A G T G       C T G   G C   C   C

181 CGGTTACCTTCCCGGGTGCTCTGGTTCCGGGTGGCGTTGCAGACGCAGCTGCTGCGTACA  240
     A     T   G     G T   A G T   T   A C T

241 AAGCGGCAAAGGCAGGTGCGGGTCTGGGCGGGGTACCAGGTGTTGGCGGTCTGGGTGTAT  300
      T T   T C T G T T T C   A   T CT A A G

301 CTGCTGGCGCAGTTGTTCCGCAGCCGGGTGCAGGTGTAAAACCGGGCAAAGTTCCAGGTG  360
      A T G G   T   T A C A G G T G   G G

361 TTGGTCTGCCGGGCGTATACCCGGGTGGTGTTCTGCCGGGCGCGCGTTTCCCAGGTGTTG  420
     G G   A T     A   C G C A A T G   C   G

421 GTGTACTGCCGGGCGTTCCGACCGGTGCAGGTGTTAAACCGAAGGCACCAGGTGTAGGCG  480
     G G C T A   C T A   A   G C   T       T

481 GCGCGTTCGCGGGTATCCCGGGTGTTGGCCCGTTCGGTGGTCCGCAGCCAGGCGTTCCGC  540
     A T T T A   A A   A C T G A   A T A C A

541 TGGGTTACCCGATCAAAGCGCCGAAGCTTCCAGGTGGCTACGGTCTGCCGTACACCACCG  600
      G T C   G C C   G T     T A   C     A
```

FIG. 6(1)

```
601  GTAAACTGCCGTACGGCTACGGTCCGGGTGGCGTAGCAGGTGCTGCGGGTAAAGCAGGCT  660
       G       C T    T G C A A G T    A    C G T T

661  ACCCAACCGGTACTGGTGTTGGTCCGCAGGCTGCTGCGGCAGCTGCGGCGAAGGCAGCAG  720
           A G A G    C C    A A A    G A T A

721  CAAAATTCGGCGCGGGTGCAGCGGGTGTTCTGCCGGGCGTAGGTGGTGCTGGCGTTCCGG  780
       G    T T A    C A C C T T T A G    T    T

781  GTGTTCCAGGTGCGATCCCGGGCATCGGTGGTATCGCAGGCGTAGGTACTCCGGCGGCCG  840
       C G T G A T T A T A C          T G    A T A

841  CTGCGGCTGCGGCAGCTGCGGCGAAAGCAGCTAAATACGGTGCGGCAGCAGGCCTGGTTC  900
       A    A    A C T G    C G T A T T     T A G

901  CGGGTGGTCCAGGCTTCGGTCCGGGTGTTGTAGGCGTTCCGGGTGCTGGTGTTCCGGGCG  960
       T    G       T C    A A T T C A A    C    A T

961  TAGGTGTTCCAGGTGCGGGCATCCCGGTTGTACCGGGTGCAGGTATCCCGGGCGCTGCGG  1020
       T    C    A T G T A    C A    T G    A T

1021 TTCCAGGTGTTGTATCCCCGGAAGCGGCAGCTAAGGCTGCTGCGAAAGCTGCGAAATACG  1080
           G    G A A    A T       A    A G A C

1081 GAGCTCGTCCGGGCGTTGGTGTTGGTGGCATCCCGACCTACGGTGTAGGTGCAGGCGGTT  1140
       G CA G C A C A    A    T T T    G T A T G C

1141 TCCCAGGTTTCGGCGTTGGTGTTGGTGGCATCCCGGGTGTAGCTGGTGTTCCGTCTGTTG  1200
       T C C T T C A C A T    T A C A    C TAG  C

1201 GTGGCGTACCGGGTGTTGGTGGCGTTCCAGGTGTAGGTATCTCCCCGGAAGCGCAGGCAG  1260
       A T T C A C A T C G A T C T    C    T
```

FIG. 6(2)

```
1261  CTGCGGCAGCTAAAGCAGCGAAGTACGGCGTTGGTACTCCGGCGGCAGCAGCTGCTAAAG  1320
        A  T  C  C  G  T  C        A  G  G  C  A  A  T

1321  CAGCGGCTAAAGCAGCGCAGTTCGGACTAGTTCCGGGCGTAGGTGTTGCGCCAGGTGTTG  1380
         C  C     C  C     T  GT        T  T  C  C  G  T  T  A

1381  GCGTAGCACCGGGTGTTGGTGTTGCTCCGGGCGTAGGTCTGGCACCGGGTGTTGGCGTTG  1440
         G  T  T     C     G     T  A  T  CT     T  T  A        G

1441  CACCAGGTGTAGGTGTTGCGCCGGGCGTTGGTGTAGCACCGGGTATCGGTCCGGGTGGCG  1500
        T  T  A  T     G  T  T        C  G  T  C  C  T  C  T     A

1501  TTGCGGCTGCTGCGAAATCTGCTGCGAAGGTTGCTGCGAAAGCGCAGCTGCGTGCAGCAG  1560
           A     A  A     C     C     G     C     C     C  A  T

1561  CTGGTCTGGGTGCGGGCATCCCAGGTCTGGGTGTAGGTGTTGGTGTTCCGGGCCTGGGTG  1620
           G  T     T        T  A  T  A  T     C  C  C  T  A  T  A

1621  TAGGTGCAGGGGTACCGGGCCTGGGTGTTGGTGCAGGCGTTCCGGGTTTCGGTGCTGGCG  1680
        T     T  T  T  T  A  T  A        T  T     T  C     G  A  T

1681  CGGACGAAGGTGTACGTCGTTCCCTGTCTCCAGAACTGCGTGAAGGTGACCCGTCCTCTT  1740
         A  T  G  A  TA G  GAG      C  T  G  CA G     A  T  C     C

1741  CCCAGCACCTGCCGTCTACCCCGTCCTCTCCACGTGTTCCGGGCGCGCTGGCTGCTGCGA  1800
         T        C  CAGC     C  A  A  CA G  A  T  A  C        C  T

1801  AAGCGGCGAAATACGGTGCAGCGGTTCCGGGTGTACTGGGCGGTCTGGGTGCTCTGGGCG  1860
           A  C     T  A     A  G  T  G  C  T  A  G  C  G     C  T

1861  GTGTTGGTATCCCGGGCGGTGTTGTAGGTGCAGGCCCAGCTGCAGCTGCTGCTGCGGCAA  1920
         A  A  C     A        G  G  A  C  A  C  C  C  C     C  A  C
```

FIG. 6(3)

```
1921  AGGCAGCGGCGAAAGCAGCTCAGTTCGGTCTGGTTGGTGCAGCAGGTCTGGGCGGTCTGG  1980
       A  T  T  C     C  C     T  C  A  G  A  C  T  G  C  A  A  C

1981  GTGTTGGCGGTCTGGGTGTACCGGGCGTTGGTGGTCTGGGTGGCATCCCGCCGGCGGCGG  2040
       A  C  A  G  T  A  T  A  T     G  C  T  A  T  A  T  A  T  A

2041  CAGCTAAAGCGGCTAAATACGGTGCAGCAGGTCTGGGTGGCGTTCTGGGTGGTGCTGGTC  2100
       C        A                 T  T  C  T  A  T  C  A  G     C  G

2101  AGTTCCCACTGGGCGGTGTAGCGGCACGTCCGGGTTTCGGTCTGTCCCCGATCTTCCCAG  2160
                T  A  A  G  A   A  A  T  C     AT    T  C  T

2161  GCGGTGCATGCCTGGGTAAAGCTTGCGGCCGTAAACGTAAATAATGATAG  2210
       T  G  C        G        T     G  GA A    G
```

FIG. 6(4)

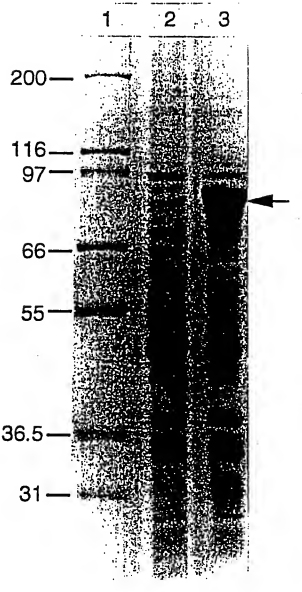

FIG. 7

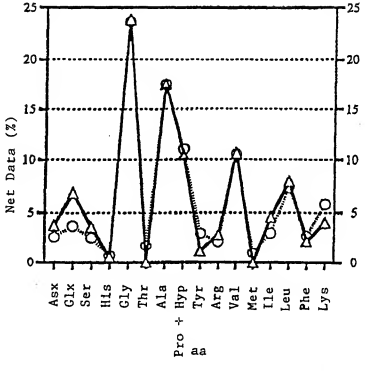
Net Data (%)
25
20
15
10
5
0
Asx
Glx
Ser
His
Gly
Thr
Ala
Hyp
Tyr
Arg
Val
Met
Ile
Leu
Phe
Lys
Pro + aa

FIG. 8

pND211 (4045 bp)
EcoR1
TTCACTGGCC GTCGTTTTAC AACGTCGTGA CTGGGAAAAC CCTGGCGTTA
CCCAACTTAA TCGCCTTGCA GCACATCCCC CTTTCGCCAG CTGGCGTAAT
AGCGAAGAGG CCCGCACCGA TCGCCCTTCC CAACAGTTGC GCAGCCTGAA
TGGCGAATGG CGCCTGATGC GGTATTTTCT CCTTACGCAT CTGTGCGGTA
TTTCACACCG CATATGGTGC ACTCTCAGTA CAATCTGCTC TGATGCCGCA
TAGTTAAGCC AGCCCCGACA CCCGCCAACA CCCGCTGACG CGCCCTGACG
GGCTTGTCTG CTCCCGGCAT CCGCTTACAG ACAAGCTGTG ACCGTCTCCG
GGAGCTGCAT GTGTCAGAGG TTTTCACCGT CATCACCGAA ACGCGCGAGA
CGAAAGGGCC TCGTGATACG CCTATTTTTA TAGGTTAATG TCATGATAAT
AATGGTTTCT TAGACGTCAG GTGGCACTTT TCGGGGAAAT GTGCGCGGAA
CCCCTATTTG TTTATTTTTC TAAATACATT CAAATATGTA TCCGCTCATG
AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA GGAAGAGTAT
GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT
GAAGATCAGT TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG
CGGTAAGATC CTTGAGAGTT TTCGCCCCGA AGAACGTTTT CCAATGATGA
GCACTTTTAA AGTTCTGCTA TGTGGCGCGG TATTATCCCG TATTGACGCC
GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA ATGACTTGGT
TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC
TTACTTCTGA CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA
CAACATGGGG GATCATGTAA CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA
ATGAAGCCAT ACCAAACGAC GAGCGTGACA CCACGATGCC TGTAGCAATG
GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA CTCTAGCTTC
CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA
GCCGGTGAGC GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG
TAAGCCCTCC CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA
TGGATGAACG AAATAGACAG ATCGCTGAGA TAGGTGCCTC ACTGATTAAG
CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATTT
AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA
GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG
CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT
GTTTGCCGGA TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTTC
AGCAGAGCGC AGATACCAAA TACTGTTCTT CTAGTGTAGC CGTAGTTAGG
CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG
TTGGACTCAA GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC
GGGGGGTTCG TGCACACAGC CCAGCTTGGA GCGAACGACC TACACCGAAC
TGAGATACCT ACAGCGTGAG CATTGAGAAA GCGCCACGCT TCCCGAAGGG
AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA CAGGAGAGCG
CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG

9/a

FIG. 9(1)

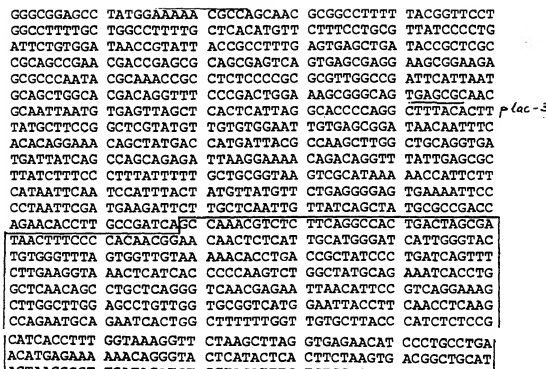
GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT
GGCCTTTTGC TGGCCTTTTG CTCACATGTT CTTTCCTGCG TTATCCCCTG
ATTCTGTGGA TAACCGTATT ACCGCCTTTG AGTGAGCTGA TACCGCTCGC
CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA
GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC
GCAATTAATG TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT
TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTGAGCGGA TAACAATTTC
ACACAGGAAA CAGCTATGAC CATGATTACG CCAAGCTTGG CTGCAGGTGA
TGATTATCAG CCAGCAGAGA TTAAGGAAAA CAGACAGGTT TATTGAGCGC
TTATCTTTCC CTTTATTTTT GCTGCGGTAA GTCGCATAAA AACCATTCTT
CATAATTCAA TCCATTTACT ATGTTATGTT CTGAGGGGAG TGAAAATTCC
CCTAATTCGA TGAAGATTCT TGCTCAATTG TTATCAGCTA TGCGCCGACC
AGAACACCTT GCCGATCAGC CAAACGTCTC TTCAGGCCAC TGACTAGCGA
TAACTTTCCC CACAACGGAA CAACTCTCAT TGCATGGGAT CATTGGGTAC
TGTGGGTTTA GTGGTTGTAA AAACACCTGA CCGCTATCCC TGATCAGTTT
CTTGAAGGTA AACTCATCAC CCCCAAGTCT GGCTATGCAG AAATCACCTG
GCTCAACAGC CTGCTCAGGG TCAACGAGAA TTAACATTCC GTCAGGAAAG
CTTGGCTTGG AGCCTGTTGG TGCGGTCATG GAATTACCTT CAACCTCAAG
CCAGAATGCA GAATCACTGG CTTTTTTGGT TGTGCTTACC CATCTCTCCG
CATCACCTTT GGTAAAGGTT CTAAGCTTAG GTGAGAACAT CCCTGCCTGA
ACATGAGAAA AAACAGGGTA CTCATACTCA CTTCTAAGTG ACGGCTGCAT
ACTAACCGCT TCATACATCT CGTAGATTTC TCTGGCGATT GAAGGGCTAA
ATTCTTCAAC GCTAACTTTG AGAATTTTTG CAAGCAATGC GGCGTTATAA
GCATTTAATG CATTGATGCC ATTAAATAAA GCACCAACGC CTGACTGCCC
CATCCCCATC TTGTCTGCGA CAGATTCCTG GGATAAGCCA AGTTCATTTT
TCTTTTTTTC ATAAATTGCT TTAAGGCGAC GTGCGTCCTC AAGCTGCTCT
TGTGTTAATG GTTTCTTTTT TGTGCTCATA CGTTAAATCT ATCACCGCAA
GGGATAAATA TCTAACACCG TGCGTGTTGA CTATTTTACC TCTGGCGGTG
ATAATGGTTG CATGTACTAA GGAGGTTGTA TGGAACAACG CATAACCCTG
AAAGATTATG CAATGCGCTT TGGGCAAACC AAGACAGCTA AAGATCTCTC
ACCTACCAAA CAATGCCCCC CTGCAAAAAA TAAATTCATA TAAAAAACAT
ACAGATAACC ATCTGCGGTG ATAAATTATC TCTGGCGGTG TTGACATAAA
TACCACTGGC GGTGATACTG AGCACATCAG CAGGACGCAC TGACCACCAT
GAAGGTGACG CTCTTAAAAA TTAAGCCCTG AAGAAGGGCA GCATTCAAAG
CAGAAGGCTT TGGGGTGTGT GATACGAAAC GAAGCATTGG GATCCTAAGG
AGGTTTAAGA TCCATGGGTT TAAACCTCCT TAGGATCCCC GGGAA EcoRI
Nco1
Bam H1
P lac -35
cI857
RBS

FIG. 9(2)

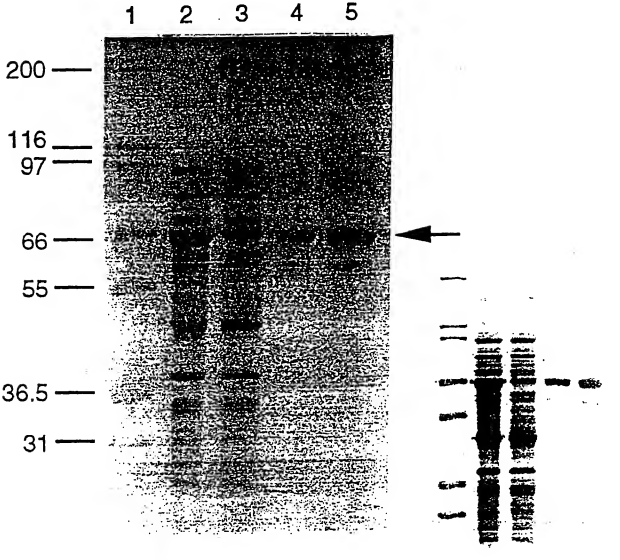

FIG. 10

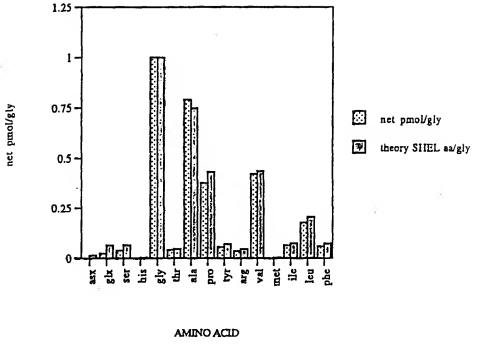
1.25
1
0.75
0.5
0.25
0
net pmol/gly
asx
glx
ser
his
gly
thr
ala
pro
tyr
arg
val
met
ile
leu
phe
net pmol/gly
theory SIIEL aa/gly
AMINO ACID

FIG. 11

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.[5] C12N 15/12, 15/62 A61K 037/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC: C12N 15/12, 15/62, Keywords as below

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AU: IPC as above

Electronic data base consulted during the international search (name of data base, and where practicable, search terms used)
DERWENT DATABASES: CHEM ABS/WPAT/BIOT/ Keywords: Tropoelastin, Elastin, T E
C12N 015/IC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	Critical reviews in eukaryotic gene expression, CRC Press Inc 1990 Volume 1 No. 3 pages 145-156, Rosenbloom et al. "Elastin Genes and Regulation of Their expression"	C1-26
Y	Archives of Biochemistry and Biophysics Volume 280, No. 1, July, pages 80-86, 1990 Indik et al. "Production of Recombinant Human Tropoelastin: Characterization and Demonstration of Immunologic and chemotatic Activity"	C1-26

[X] Further documents are listed in the continuation of Box C.

[ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
?	25 March 1994 (25.03.94)
Name and mailing address of the ISA/AU	Authorized officer
AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA	T RICHARDS
Facsimile No. 06 2853929	Telephone No. (06) 2832445

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate of the relevant passages	Relevant to Claim No.
Y	Biochemistry Volume 26, No. 6 24 March, 1987 Bressan et al. "Repeating Structure of Chick Tropoelastin Revealed by Complementary DNA cloning", pages 1497 to 1502	C1-26
Y	Biotechnology progress Volume 6, 1990 pages 198-202 Capello et al. "Genetic engineering of structural protein Polymers"	C1-11
A	Annals of the New York Academy of Sciences Volume 624 1991 pages 116-36 Rosenbloom et al. "Regulation of Elastin Gene expression"	C1-26
A	Biotechnol Prog, Volume 8 1992 pages 347-352 McPherson et al. "Production and Purification of Recombinant Elastomeric Polypeptide, G-(VPGVG)$_{19}$-VPGV from E. coli."	C1-26
A	The Journal of Biological Chemistry, Volume 265 No. 16 1990, Kahari et al. "Deletion Analysis of 5' Flanking Region of Human Elastin Gene" pages 9485-9490	C1-26
A	The Journal of Biological Chemistry, Volume 262 No. 12, 1987, Ragu et al. "Primary Structure of Bovine Elastin a, b and c Deduced from the sequences of cDNA Clones" pages 5755-5762	C1-26